

Derivation of a UK preference-based value set for the Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS) to allow estimation of Mental Well-being Adjusted Life Years (MWALYs)

Abstract:

Background: The Mental Well-being Adjusted Life Year (MWALY) is an alternative outcome measure to the quality-adjusted life year (QALY) in economic evaluations of interventions aimed at improving mental well-being. However, there is a lack of preference-based mental well-being instruments for capturing population mental well-being preferences.

Objectives: To derive a UK preference-based value set for the Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS).

Methods: 225 participants that were interviewed between December 2020 and August 2021 completed 10 composite time trade-off (C-TTO) and 10 discrete choice experiment (DCE) interviewer-administered exercises. Heteroskedastic Tobit and conditional logit models were used to model C-TTO and DCE responses respectively. The DCE utility values were rescaled to a C-TTO comparable scale through anchoring and mapping. An inverse variance weighting hybrid model (IVWHM) was used to derive weighted-average coefficients from the modelled C-TTO and DCE coefficients. Model performance was assessed using statistical diagnostics.

Results: The valuation responses confirmed the feasibility and face validity of the C-TTO and DCE techniques. Apart from the main effects models, statistically significant associations were estimated between the predicted C-TTO value and participants' SWEMWBS scores, gender, ethnicities, education levels, and the interaction terms between age and useful feeling. The IVWHM was the most optimal model with the fewest logically inconsistent coefficients and the lowest pooled standard errors. The utility values generated by the rescaled DCE models and the IVWHM were generally higher than those of the C-TTO model. The predictive ability of the two DCE rescaling methods was similar according to the mean absolute deviation and root mean square deviation statistics.

Conclusions: This study has produced the first preference-based value set for a measure of mental well-being. The IVWHM provided a desirable blend of both C-TTO and DCE models. The value set derived by this hybrid approach

25 can be used for cost-utility analyses of mental well-being interventions.

26 *Keywords:* Preference elicitation, Short Warwick-Edinburgh Mental Well-being Scale, Composite time trade-off,

27 Discrete Choice Experiment, Inverse variance weighting hybrid, Mental well-being, United Kingdom.

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45 **Highlights:**

- 46 - This is the first attempt to derive a preference-based value set for a generic mental well-being instrument.
- 47 - The valuation responses supported the face validity and feasibility for the valuation of the SWEMWBS.
- 48 - The Inverse Variance Weighting hybrid model is firstly applied in preference elicitation research.
- 49 - Units of QALY gain and MWALY gain have different implications for economic evaluations of mental well-
- 50 being interventions.
- 51 - This study tests the use of remote-control function in videoconferencing during the interview.

52

53

54

55

56

57

58

59

60

61

62

63

64

1. Introduction

The EQ-5D (Brooks & Group, 1996) has long been recommended by health technology agencies for estimating quality-adjusted life years (QALYs) for cost-utility analysis purposes. However, the EQ-5D is limited by its focus on physical health outcomes. Four of the five EQ-5D dimensions relate to physical health (mobility, self-care, usual activities, pain/discomfort) and only one dimension relates to mental health (anxiety/depression). Previous studies have questioned whether the EQ-5D can capture mental health outcome changes for individuals with delusional or bipolar I disorder, and other psychotic disorders (Brazier, 2010; Saarni *et al.*, 2010). The low proportion of mental health-related dimensions covered by the EQ-5D hinders its sensitivity in detecting outcome changes related to mental health.

Due to the potential underestimation of mental health related intervention benefits captured by QALYs in economic evaluation (Brazier, 2010; Saarni *et al.*, 2010; Shah *et al.*, 2017), there has been an increasing number of patient-reported outcome measures developed with broader coverage of mental health items. For example, the development of the Oxford CAPabilities questionnaire-Mental Health (OxCAP-MH) allows the measurement of different dimensions of capability well-being (e.g. daily activities, social networks, freedom of expression, self-determination, etc.). The OxCAP-MH preference study provided insights regarding the impact of mental ill-health related heterogeneity on the capability-related mental health preferences of the general population (Helter *et al.*, 2022). Also, the Mental Well-being Adjusted Life Year (MWALY) has been suggested as an alternative outcome measure to capture the benefits associated with improvements in population mental well-being (Johnson *et al.*, 2016). The Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS) (Supplementary Information 1) is a 7-item questionnaire widely recommended for measuring mental well-being, currently translated into 30 languages and used by agencies in England, Scotland, and Wales in the UK for monitoring population mental well-being (Diana Bardsley *et al.*, 2017; Parkinson, 2007; Scottish Government, 2018). It is a generic mental well-being instrument that is suitable for completion by the general population. Each of the seven items consists of response levels across a five-point Likert scale ranging from “none of the time” to “all of the time”. The minimum and maximum scores for the SWEWMWBS are 7 and 35, respectively. A higher score reflects a better mental well-

90 being of an individual in the general population, and vice versa. A SWEMWBS score of >18-20 indicates possible
91 mild depression, whereas a score of ≤ 18 indicates probable clinical depression (Shah *et al.*, 2021). The SWEMWBS
92 is widely used to measure the impact of interventions to promote mental health and is responsiveness to change
93 (Shah *et al.*, 2018). It has been psychometrically validated in different population groups across the world and
94 SWEMWBS responses have been shown to lack floor and ceiling effects, and to be uni-dimensional with high
95 internal consistency, good construct validity, good test-retest reliability and high face validity (Anthony *et al.*, 2022;
96 Bass *et al.*, 2016; Haver *et al.*, 2015; Koushede *et al.*, 2019; Ng Fat *et al.*, 2017; Ng *et al.*, 2014; Rogers *et al.*, 2018;
97 Shah *et al.*, 2021; Vaingankar *et al.*, 2017).

98 The SWEMWBS includes hedonic or (feeling) items as well as eudemonic (functioning) items and therefore has a
99 slightly different construct from the OxCAP-MH, which more closely maps onto capabilities. MWALYs based on
100 SWEMWBS therefore arguably work better as an outcome measure for public mental health interventions, which
101 usually aim to improve both feeling and functioning. Also, the SWEMWBS was designed to capture mental well-
102 being and is therefore more likely to capture mental health than generic health-related quality of life measures such
103 as the EQ-5D. With its focus on the positive, it is also designed to capture a broader range of mental health outcomes
104 than the EQ-5D. Quantitative evidence shows the distinction in construct measurement between the EQ-5D and
105 WEMWBS (i.e. a longer version of the SWEMWBS with 14 items). Johnson *et al.* (2016) analysed response data
106 from the Coventry Household Survey to explore the mapping relationship between EQ-5D-3L and WEMWBS.
107 More than 70% of the participants who scored a maximum of 1 in the EQ-5D-3L reported WEMWBS scores
108 between 14 and 70, with a mean score of around 53.9. In linear models, WEMWBS scores explained less than 15%
109 of the variability in the EQ-5D-3L scores. These results suggest that they are not measuring the same aspects of
110 health or well-being. Depending on the research questions and types of interventions evaluated, the EQ-5D is likely
111 to be a better choice for evaluating physical health related interventions and the SWEMWBS is likely to perform
112 better when evaluating interventions aimed at improving mental health and well-being. In this context, the
113 derivation of a preference-based value set for the SWEMWBS fits the purpose of MWALY estimation.

114 A previous think-aloud study investigated the patterns and problems in using composite time trade-off (C-TTO)

115 and discrete choice experiment (DCE) for the valuation of the SWEMWBS (Yiu *et al.*, 2022). C-TTO is a choice-
116 based preference elicitation method developed based on the concept of opportunity cost (John Brazier *et al.*, 2017).
117 It allows respondents to choose between living in a hypothetical state A for a particular life span and in a better
118 hypothetical state B but with a shorter life span. The indifference point between these two states indicated by the
119 respondent represents the number of years of life the respondent is willing to give up, to avoid being in the
120 hypothetical state A. If the respondent is indifferent between zero years of life and 10 years of living in a
121 hypothetical mental well-being state, this response is interpreted as the respondent being willing to give up all years
122 of mental well-being to avoid being in that designated mental well-being state. The state “death” is a proxy for no
123 years of mental well-being. As the C-TTO consists of both better-than-dead and worse-than-dead scenarios, the
124 preference values generated by the C-TTO responses lie on a scale that encompasses full health (well-being) and
125 death. DCE is a stated preference method developed under random utility theory (Lancsar & Louviere, 2008). The
126 value assigned to a hypothetical scenario is inferred by the respondent’s choice among multiple hypothetical
127 alternatives described by various attributes and levels. However, unlike the C-TTO method, utility values derived
128 from DCE responses are estimated on a latent scale. Rescaling methods should be used to anchor DCE utility values
129 to a full health and death scale for use in economic evaluations. As the qualitative study (Yiu *et al.*, 2022) showed
130 that the application of C-TTO and DCE exercises for the valuation of the SWEMWBS was suitable for capturing
131 individual preferences towards mental well-being, this study aims to derive a UK preference-based value set for the
132 SWEMWBS.

133

134 **2. Methods**

135 Structured interviews with the presence of an interviewer (HHEY) were administered in a Computer-Assisted
136 Personal Interview (CAPI) setting. The EQ-PVT platform (i.e. a replica of the EQ-VT 2.1) developed by the
137 EuroQol Group (Stolk *et al.*, 2019) was used to record participant responses to 10 C-TTO (i.e. conventional TTO
138 for the valuation of mental well-being states considered better than death and a lead-time TTO for states considered
139 worse than death) and 10 DCE tasks (examples shown in Figures 1 and 2). Following COVID-19 related restrictions

in place during the study, interviews were held using an online meeting platform, Microsoft Teams. Individuals who were members of the general UK population, had the self-asserted right to vote in the U.K. and aged 18 or above were eligible for participation. This research was approved by the Biomedical and Scientific Research Ethics Committee at the University of Warwick (Reference: BSREC.44/19-20).

Participants were recruited from convenience or snowball sampling through advertising on social media (Facebook, Instagram, Reddit and Twitter), personal networks, and University platforms for reaching members of the general population across England, Scotland, Wales and Northern Ireland. Social media covered a wide variety of advertising placements, including mobile app news feed, Instant Article, Instagram Stories, Instagram feed, and Mobile in-stream video, etc., enhancing recruitment of participants from unknown contacts. A £25 Amazon shopping voucher was given to each of the 10 randomly drawn winners.

2.1. Experimental design

2.1.1. DCE

A Bayesian D-efficient design, with priors obtained from the previous qualitative study, was used to generate choice tasks using Ngene (Yiu *et al.*, 2022). Ngene software imposes attribute level balance in the design (ChoiceMetrics, 2018). Fifty generated choice tasks were divided into five blocks (syntax reported in Supplementary Information 2). Each participant was randomly assigned to one of the blocks. The sample size calculation was informed by an approximate formula (Louviere *et al.*, 2000):

$$N \geq \frac{1-p}{Tp(a^2)} \left[\varphi^{-1} \left(\frac{1+\alpha}{2} \right) \right]^2$$

For a pairwise DCE with T (choice tasks) = 10, p (expected choice proportion) = 0.5, a (accuracy) = 0.1, α (confidence level) = 0.95, φ^{-1} = inverse normal cumulative distribution, $N \geq 39$. As there were 5 blocks, the required minimum sample size $N \geq 195$ (39×5). The final design with the lowest D-error is provided in Supplementary Information 3.

2.1.2. C-TTO

A blocked design used the level-balance optimisation criterion provided by the EuroQol Group (Oppe & Van Hout, 2017). This criterion suggests that the number of appearances of each level-domain combination is counted for checking whether each level of one attribute appears the same number of times. The value of the level balance check was calculated using the following formula:

Value of the level balance check

$$= \sqrt{\text{sum of squares of the differences between the presence of levels per attribute}}$$

The whole algorithm was run over 10,000 iterations to generate a subset with the lowest value in level balance check (i.e. the best achievement on level balance). R software was used to generate seven blocks of choice tasks. Each participant completed two compulsory mental well-being states [lowest mental well-being state (111111) and one of the closer to full mental well-being states (455555, 545555, 554555, 555455, 555545 or 555554)] plus eight randomly generated mental well-being states. The seven generated blocks, which achieved the lowest value for level balance compared to other generated designs, are provided in Supplementary Information 4.

2.2. Data analysis

Modelling analyses were performed in Stata 16.0 to estimate preference scores for all 78,125 SWEMWBS mental well-being states.

2.2.1. Heteroskedastic Tobit model (HTM) for the C-TTO

The C-TTO task consisted of two parts: conventional TTO for the valuation of better-than-dead states and lead-time TTO for the valuation of worse-than-dead states. As the number of years of full mental well-being in life A lay between 0 and 10 and the number of years of the lower than full-mental well-being state in life B was set at 10 years in the better-than-dead scenario, the generated value was bounded between 0 and 1. The value generated from the worse-than-dead scenario lay between -1 and 0, given that the ratio of lead-time to duration of the mental well-being state was 1:1. The C-TTO value generated by both better-than-dead and worse-than-dead scenarios therefore ranged between -1 and 1, assuming that the participant was able to reach the indifference point between life A and

life B. However, our previous SWEMWBS think-aloud valuation study discovered a few observations with C-TTO values < -1 and > 1 , due to participants' failure to achieve the indifference point for tasks (Yiu *et al.*, 2022). A limited dependent variable left censored at -1 and right censored at 1 was therefore used to model the C-TTO data. The main effects specification was as follows:

$$Y_{C-TTO} = \beta_0 + f(ITEMi_{Lj}\beta) + \varepsilon$$

Y_{C-TTO} = A dependent variable, representing the C-TTO value; β_0 = A constant term; $ITEMi_{Lj}$ = A group of dummy variables $ITEMi$ for a specific level Lj ($i = 1, 2, 3, \dots, 7$ and $j = 1, 2, 3, 4$); reference category = level 5 (the highest mental well-being level); β = A group of coefficients indicating utility changes from the highest mental well-being state; ε = An error term.

As the standard deviation of the C-TTO response data generally increased with the lower mental well-being state (Supplementary Information 5), an HTM was used, assuming normally distributed residuals. Our previous qualitative study suggested that participants with experience of mental illness tended to give lower C-TTO values to mental well-being states (Yiu *et al.*, 2022). To explore this, the SWEMWBS score obtained by participants' self-completion of the SWEMWBS was added as an individual-specific covariate to the model. The study also suggested a potential association between age and the item "I've been feeling useful". Interaction terms between age and this item were therefore added. Even though the relationships between other individual-specific covariates (i.e. gender, ethnicity and education level) and C-TTO values were not documented in the previous study, these individual-specific covariates were added to explore potential group-based effects.

2.2.2. Conditional Logit model (CLM) for the DCE

The CLM proposed by McFadden was used to model the data (McFadden, 1973). An individual-specific covariate can only be interacted with constants or attributes, prohibiting their use as controlled variables (Greene, 2003). Deterministic heterogeneity was explored by the interaction terms between age and the item "I've been feeling useful", as informed by the results of the qualitative study (Yiu *et al.*, 2022).

Utility values were estimated on a latent scale and rescaled to a C-TTO comparable scale indexed at 0 and 1 using

two methods (Bahrampour *et al.*, 2020; Rowen *et al.*, 2015). The first method anchored the DCE value of the lowest mental well-being state (i.e. 1111111) at the C-TTO value of the lowest mental well-being state (i.e. 1111111) ($DCE_{\text{anchoring}}$), assuming the DCE unscaled value is linearly proportional to the DCE rescaled value by a factor of α , i.e. *DCE rescaled value for a state i = α DCE unscaled value for a state i + γ* , where γ = a constant term. The *DCE rescaled value* was replaced with the C-TTO value generated by the HTM for the lowest mental well-being state. The *DCE unscaled value* was replaced with the DCE latent value generated by the CLM for the lowest mental well-being state. By definition, the C-TTO value of the highest mental well-being was 1. The unscaled DCE value of the highest mental well-being state was normalised at 0. The constant term γ was therefore set at 1, to ensure identical values generated by the rescaled DCE model and the C-TTO model for both the highest and lowest mental well-being states.

The second method mapped DCE values onto C-TTO values via OLS regression (DCE_{mapping}), to derive utility values for all C-TTO states based on the latent DCE values, i.e.

$$TTO = f(DCE) + \varepsilon \quad \text{where}$$

TTO = The mean C-TTO value of each of the 64 SWEMWBS states valued by the participants during the completion of C-TTO tasks;

DCE = The latent DCE utility values for each of these 64 states;

ε = An error term.

Based on the plot (Supplementary Information 6) between the mean C-TTO values and the latent DCE values for the 64 states, an overall linear association between C-TTO values and latent DCE values was assumed.

2.2.3. Inverse Variance Weighting hybrid model (IVWHM) for both the C-TTO and DCE

This approach derived weighted-average coefficients for the attribute levels from the modelled C-TTO and rescaled DCE coefficients (Lee *et al.*, 2016). The weights favoured the coefficients with lower standard errors (SEs) (i.e. lower uncertainties), to ensure the generation of more reliable pooled coefficients. The pooled coefficient of an

attribute level was calculated by assuming that the C-TTO and rescaled DCE coefficients were independent and normally distributed. In other words, each pooled coefficient was the sum of the C-TTO and rescaled DCE coefficient, each weighted by the inverse of their variance, and the variance of the pooled coefficient was the weighted summation of the C-TTO and rescaled DCE variances.

The pooled coefficients and the pooled SEs were calculated as follows:

$$Pooled\ coefficient_{ij} = TTO_{ij} * \left[\frac{\frac{1}{(S_{TTO_{ij}})^2}}{\frac{1}{(S_{TTO_{ij}})^2} + \frac{1}{(S_{DCE_{ij}})^2}} \right] + DCE_{ij} * \left[\frac{\frac{1}{(S_{DCE_{ij}})^2}}{\frac{1}{(S_{TTO_{ij}})^2} + \frac{1}{(S_{DCE_{ij}})^2}} \right]$$

$$Pooled\ Standard\ error_{ij} = \sqrt{\left[S_{TTO_{ij}} * \frac{\frac{1}{(S_{TTO_{ij}})^2}}{\frac{1}{(S_{TTO_{ij}})^2} + \frac{1}{(S_{DCE_{ij}})^2}} \right]^2 + \left[S_{DCE_{ij}} * \frac{\frac{1}{(S_{DCE_{ij}})^2}}{\frac{1}{(S_{TTO_{ij}})^2} + \frac{1}{(S_{DCE_{ij}})^2}} \right]^2}$$

, where

$Pooled\ coefficient_{ij}$ = the weighted-average pooled coefficient for the attribute i with level j ; TTO_{ij} = the coefficient for the attribute i with level j derived from the C-TTO model; DCE_{ij} = the rescaled coefficient for the attribute i with level j derived from the DCE model; $S_{TTO_{ij}}$ = the SE for the attribute i with level j derived from the C-TTO model; $S_{DCE_{ij}}$ = the rescaled SE for the attribute i with level j derived from the DCE model.

2.2.4. Model analysis

Supplementary Information 7 provides a description of all explanatory variables included in the modelling of C-TTO and/or DCE data. The criteria for assessing the model performance were as follows:

- Logical consistency of the estimators: A lower (higher) level of an item should theoretically generate a higher (lower) utility decrement.
- Statistical significance of the estimators.

- Goodness of fit: This was assessed by the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) to examine the efficiency of models, with an attempt to strike balance between the sophistication (complexity) and the simplicity of the model. Also, a Wald test was conducted to explore the addition of covariates on the model fit improvement.
- Parsimony: A simple model with optimal number of predictors (without overfitting) was sufficient to explain the model well.
- Mean absolute deviation (MAD) and root mean square deviation (RMSD) (formulae in Supplementary Information 8): The two DCE rescaling methods were compared to investigate their ability to predict the observed C-TTO values for the 64 states valued by the participants (Rowen *et al.*, 2015).

2.3. Interview process

All interviews were audio and screen recorded. The procedural elements within each interview are highlighted below.

- (1) The interviewer introduced the study.
- (2) The participant signed a consent form.
- (3) The participant completed the SWEMWBS to describe their own mental health status over the past two weeks, and sociodemographic questions.
- (4) The C-TTO exercise: The participant selected between two versions of warm-up example regarding the imagination of mental well-being scenarios (Supplementary Information 9). The participant was then guided with step-by-step explanations about the trade-off process of both better-than-death and worse-than-death scenarios, and the way of using the remote-control function in Microsoft Teams to identify the preferred life on their own. Three extra practice states were then followed: high (4554545), low (2111131) and intermediate (3313432) states. Next, the participant completed 10 tasks. Afterwards, the rank ordering inferred by their valuations was displayed on the Feedback Module (Figure 3). The participant was asked to flag any disagreements of the results without altering the problematic valuations. The flagged valuations should be removed from the data (Shah *et al.*, 2014). Finally,

275 debriefing questions (Supplementary Information 10) were asked.

276 (5) The DCE exercise: The participant chose the preferred option among 10 pairs of mental well-being states.
277 The paired comparisons and the left-right order of the two states were randomised by the EQ-PVT. Debriefing
278 questions (Supplementary Information 10) were asked.

279 (6) Overall debriefing questions (Supplementary Information 10) were asked.

280 **2.4. Sensitivity analysis**

281 For the modelling of the C-TTO data, the estimated parameters were investigated when the values of the flagged
282 C-TTO states in the Feedback Module were included. For the DCE data, responses with potential strategic/repeated
283 pattern were excluded.

284 A post-stratification adjustment on the individual-level data generated weights to correct covariate imbalance in our
285 sample relative to the general population. These weights were applied to regressions for the DCE and TTO, to
286 reflect the effect of pre-specified demographic distribution on the preference elicitation results.

287

288 **3. Results**

289 All interviews were conducted between December 2020 and August 2021. In total, 227 participants attended at their
290 scheduled time. Two withdrawals during the completion of C-TTO exercise were due to exhaustion or technical
291 difficulties. The number of completed interviews was 225. The mean completion time was 60 minutes per interview.
292 Table 1 describes the characteristics of the participants.

293 Relative to the general UK population, these 225 participants consisted of a relatively greater proportion of females
294 (73.33% v.s. 50.59% in the UK) than males (25.33% v.s. 49.41% in the UK) (Office for National Statistics). The
295 median age (51 years) was also higher than the medians age (40.4 years) in the UK (Office for National Statistics).
296 Even though the sample covered people with diverse education backgrounds, people with low or no qualifications
297 were underrepresented. Nevertheless, the ethnicity distribution of the study sample was very similar to that of the

298 UK population, i.e. 87% of Whites, 7% of Asian/Asian British, 3% of Black/African/ Caribbean/Black British, 2%
299 of Mixed and 1% of others (Office for National Statistics). Around 18% of our sample had a SWEMWBS score
300 ≤ 20 , indicating possible mild or clinical depression. Around 10% of our sample had a SWEMWBS score between
301 31-35, indicating a very high level of mental well-being. The mean SWEMWBS score of our sample was 24.8,
302 which was similar to the mean SWEMWBS score of 23.5 for the general population in the UK (Ng Fat *et al.*, 2017).
303 The mental health status of our sample was therefore representative of the general population.

304 The response statistics of the C-TTO and DCE debriefing questions supported the feasibility or practicality of the
305 application of the two valuation techniques for the valuation of mental well-being states (Supplementary
306 Information 11).

307 **3.1. The C-TTO models**

308 The 147 flagged responses in the Feedback Module were deleted, resulting in 2103 valid responses. The distribution
309 of the C-TTO values is plotted in Figure 4. The distribution was left-skewed, as more C-TTO values were clustered
310 at the positive end. The most common C-TTO value was 1, as indicated by the peak of distribution. The second
311 most frequent value was 0.5, followed by the value of 0.8. There were 123 responses with negative C-TTO values,
312 contributing only 5.85% of the total responses. States with higher level-sum score were generally associated with
313 higher mean C-TTO values (Supplementary Information 12), supporting the face validity of the C-TTO technique.

314 The main effects model is labelled as Model 1A (Table 2). The sign of coefficients was generally in negative
315 directions, indicating mental well-being disutility compared to the reference level “all of the time”. The number of
316 statistically insignificant main effects parameters at the 5% level was nine. There were 11 potentially logical
317 inconsistent coefficients (highlighted in bold). For example, the utility decrement for “optimistic4” was higher than
318 that for “optimistic3”.

319 Individual-specific covariates related to the participants’ SWEMWBS score, gender, ethnicity and education level
320 were added in Model 1B (Supplementary Information 13). Model 1B had fewer logically inconsistent coefficients,
321 better explanatory power, and goodness of fit. The covariates of age and the interaction terms between age and the

item “I’ve been feeling useful” were added to Model 1B to create Model 1C (Supplementary Information 13). However, Model 1C was not preferred to Model 1B based on the principle of parsimony.

3.2. The DCE models

After deleting four missing answers, 2246 valid DCE responses remained. Generally, participants preferred option A (B) when the difference in level-sum score was positive (negative) (Supplementary Information 14), supporting the face validity of the DCE.

The main effects model is labelled as Model 2A in Table 2. The marginal utility for all main effects coefficients were negative, except for the coefficients of “relaxed4” and “thinkingclearly4” (highlighted in bold). Seven coefficients were statistically insignificant. The coefficient for the constant term was negative and significant at the 5% level, suggesting a left-right bias. Participants tended to choose the right-hand alternative.

In Model 2B, age was interacted with the attribute “I’ve been feeling useful” (Supplementary Information 15). However, these coefficients were all statistically insignificant.

3.2.1. $DCE_{anchoring}$

The rescaled coefficients in Model 2A are reported in Table 2. Although Model 1B was preferred to Model 1A, the main effects coefficients were very similar. To ensure the compatibility of the model variables, Model 1A for the C-TTO data alongside Model 2A for the DCE were used for value set calculation and anchoring of the DCE values. The C-TTO utility value generated by Model 1A for state 1111111 = $1 - 0.153 - 0.144 - 0.168 - 0.097 - 0.204 - 0.169 - 0.165 = -0.1$. The DCE unscaled utility value generated by Model 2A for state 1111111 = $0 - 1.502 - 1.443 - 1.369 - 1.392 - 1.261 - 2.239 - 1.254 = -10.46$. Hence, $\alpha =$

$$\frac{C-TTO \text{ utility value for the state } 1111111 - 1}{DCE \text{ unscaled utility value for the state } 1111111} = 0.105$$

The rescaled factor $\alpha = 0.105$ was applied to rescale any DCE utility value for a particular state.

3.2.2. $DCE_{mapping}$

Supplementary Information 16 shows the result generated by regressing the mean C-TTO value on the DCE

unscaled values for the 64 SWEMWBS states. The coefficients of unscaled DCE utility values and the constant term were statistically significant at the 1% level. These coefficients were applied to calculate the rescaled DCE values from the unscaled DCE values. The weight of utility change from level 5 for all attribute levels were derived (Table 3).

3.3. *The IVW hybrid model (IVWHM)*

For Model 3 in Table 2, there was only one potentially logical inconsistent coefficient (thinkingclearly4). The pooled coefficients were derived from the weighted average of the C-TTO and rescaled DCE coefficients. The C-TTO and rescaled DCE coefficients for the attribute levels were extracted from the main effect models (i.e. Model 1A and Model 2A). Given the two DCE rescaling methods, coefficients for the $DCE_{\text{anchoring}}$ were selected as the rescaled DCE coefficients, as the SEs for the anchored coefficients were already available in Table 3. For the attribute level “optimistic1”, the C-TTO SE (0.0363) was higher than the rescaled DCE SE (0.0288). As a result, the DCE rescaled coefficient (-0.158) with relatively lower uncertainty received a higher weight. The pooled coefficient (-0.156) inclined towards the rescaled DCE coefficient. The pooled SE for the attribute levels was derived under the assumption that the C-TTO and DCE rescaled coefficients were independent and normally distributed.

3.4. *Sensitivity analysis*

The C-TTO main effect model with the inclusion of flagged states is labelled as Model 4 (Supplementary Information 17). There was no obvious difference in the model fit with and without the exclusion of flagged states concerning the number of statistically insignificant coefficients and the AIC/BIC statistics. Nevertheless, the Model 1A was still preferred to Model 4, as it was not sensible to include the values of states with incorrect rank ordering claimed by the participants in the Feedback Module.

The result of the sensitivity analysis for the DCE main effect model is labelled as Model 5 (Supplementary Information 18). After excluding 40 potentially strategic/repeated answers (BBBAABBBAA, BBBBBA AAAA, AAAAAAAAAA, ABAABABAAB), differences in model performance between Model 5 and the Model 2A were

negligible. Model 2A remains preferred to Model 5 as there was no evidence to justify the participants' disengagement in providing these responses.

As the proportions of categories of gender, age and education level of this study sample was not truly representative of the general population, these demographic variables were weighted to simulate a representative sample of the general population. The results for the weighted C-TTO main effect model (Supplementary Information 19) and weighted DCE main effect model (Supplementary Information 19) were similar and consistent with the unweighted main effect models in Table 2.

3.5. Comparison of valuation sets

The coefficients for the attribute levels generated by Model 1A (C-TTO main effect model) were comparable to those for Model 2A ($DCE_{\text{anchoring}}$) as they lay on the same utility scale. Model 2A had far fewer logically inconsistent coefficients than Model 1A, and slightly more statistically significant parameters. Only one inconsistent coefficient (thinkingclearly4) was identified in Model 3 (IVWHM), which was the lowest.

The coefficients in Model 2A had generally smaller rescaled SEs than in Model 1A. However, these comparative advantages of Model 2A over Model 1A did not necessarily suggest superiority of DCE over C-TTO for modelling mental well-being preferences. Arguably, the performance of the C-TTO model improved after adding individual covariates in Model 1B, even though it could not be compared directly to Model 2A due to different sets of explanatory variables. Compared to Model 1A and Model 2A, the pooled SEs surrounding all coefficients in Model 3 were the lowest.

Additionally, concerning the result of DCE_{mapping} , the utility value of the highest mental well-being state 5555555 generated by mapping (i.e. 0.954) was not identical to that generated by Model 1A, $DCE_{\text{anchoring}}$, and Model 3 (i.e. 1). The values for level utility changes from level 5 for each attribute generated by the DCE_{mapping} therefore could not be compared directly to the level coefficients generated by the C-TTO model, the $DCE_{\text{anchoring}}$, and the IVWHM. Nevertheless, the pattern of the utility change for the DCE_{mapping} could still be explored. As with the $DCE_{\text{anchoring}}$, there were two logically inconsistent values (relaxed4, thinkingclearly4) identified by the DCE_{mapping} . In short, the

rescaled DCE main effects model tended to perform relatively better than the C-TTO main effects model. The ranking of attributes and the largest and smallest level change from the base level were different across the C-TTO, DCE and the IVWHMs, but they were identical between the two rescaled DCE models.

The four value sets generated by the C-TTO and DCE rescaling models are displayed in Figure 5. In general, higher (lower) utility values were found for states with higher (lower) levels of mental well-being. For the C-TTO value set, the lowest mental well-being state 1111111 did not receive the lowest utility value (-0.1). The lowest utility value calculated by this model was -0.13 for states 1122111 and 1122121. The highest utility value was 1.000227 for state 555545, which was roughly identical to the utility value of 1 for state 555555. For the DCE_{anchoring} value set, the lowest utility value was -0.1 for state 1111111. There were nine states with utility values >1, indicating a utility increment from the highest mental well-being state 555555. The highest utility value was 1.014 for state 554545. The DCE_{mapping} value set was the only value set without any negative utility value. The lowest utility value was 0.075 for state 1111111, whereas the highest utility value was 0.965 for state 554545. For the IVWHM value set, the lowest utility value was -0.09 for state 1111111. There were four states with utility values >1.

When investigating the four value sets together, the value sets generated by DCE_{anchoring} and DCE_{mapping}, and the IVWHM were generally higher than that of the C-TTO model. The utility values generated by the DCE_{mapping} were generally higher than those of the DCE_{anchoring} and the IVWHM method across the range of low and intermediate mental well-being states. The utility values of DCE_{anchoring} and IVWHM were higher for some states than those generated by the DCE_{mapping} at the high-end of mental well-being. The value sets generated by DCE_{anchoring} and DCE_{mapping} correlated perfectly to each other (correlation = 1). The IVWHM value set and either of the DCE rescaling value sets (0.992) or the C-TTO value set (0.957) were highly correlated. The correlation between the C-TTO value set and either of the DCE rescaling value sets was also high (0.921). Regarding the predictive ability to the observed C-TTO values, the performance of DCE_{anchoring} and DCE_{mapping} in terms of the MAD and RMSD was similar. Also, the overall deviation between the rescaled utility values and the observed mean C-TTO values for the 64 SWEMWBS states tended to be lower. There was no prediction with rescaled utility greater than 0.2 from the observed mean C-TTO.

4. Discussion

This study applied Heteroskedastic Tobit, conditional logit and inverse variance weighting hybrid modelling approaches to model the C-TTO and DCE responses provided by 225 UK participants. The results produced the first value sets for a generic mental well-being measure. The C-TTO and DCE responses supported the face validity and feasibility for the valuation of the SWEMWBS.

When exploring the performance of main effects models, the unscaled DCE model (Model 2A) performed better than the C-TTO model (Model 1A) in terms of fewer potentially logical inconsistent coefficients and more statistically significant coefficients for the attribute levels. Unlike the C-TTO model, the main difference of the DCE model was that the generated latent utility values were not bounded within a scale indexed at 0 and 1 to be used in cost-utility analyses. To allow the estimation of MWALYs based on the DCE modelling results, the C-TTO data were used to rescale DCE utility values onto a C-TTO comparable scale. This study presented two ways to rescale the unscaled DCE coefficients into a C-TTO comparable scale. $DCE_{\text{anchoring}}$ anchored the DCE unscaled value of the lowest mental well-being state at the C-TTO value of the lowest mental well-being state. The performance of the resulting rescaled DCE coefficients was very similar to the unscaled DCE coefficients in terms of identical number of statistically significant coefficients and potentially logical consistent coefficients. Also, all rescaled DCE coefficients had lower SEs than the unscaled DCE coefficients. This suggested that the impact of statistically insignificant and potentially logical inconsistent coefficients for the C-TTO model on the robustness of rescaled DCE coefficients was negligible. DCE_{mapping} mapped DCE values onto C-TTO values by regression analysis. Similar to the results reported by Rowen *et al.* (2015), this study found that the utility values generated by the DCE_{mapping} performed slightly better than the $DCE_{\text{anchoring}}$ in terms of predicting the observed C-TTO values. However, the comparison of predictive ability of the observed C-TTO values measured by the MAD or RMSD was rough indicators, as arguably the mean observed value was no longer the best indication of central tendency because the C-TTO values were censored (Feng *et al.*, 2018).

442 Independent DCE rescaling methods are available that do not rely on data from C-TTO or other valuation
443 techniques. For example, Norman *et al.* (2014) included a survival duration attribute within their DCE choice sets
444 to explore trade-offs between quality of life and life expectancy. Stolk *et al.* (2010) attempted to anchor DCE values
445 using a dummy coefficient for dead. Respondents were asked to compare each of the two EQ-5D health profiles
446 within the DCE pair to a choice of dead. Despite these rescaling methods, there was a lack of research investigating
447 ways in which to convert DCE values onto a scale indexed at 0 and 1 without relying on data from other valuation
448 techniques when there was no dead or duration attribute available in the choice tasks.

449 Similar to the sensitivity analysis results of the Danish EQ-5D-5L valuation study (Jensen *et al.*, 2021), substantial
450 differences of model performance with and without the exclusion of the flagged C-TTO values in the Feedback
451 Module were not observed. Arguably, these could be partly due to the small number of flagged answers, as only
452 147 (6.53%) out of the 2250 responses were deleted in Model 1A. Notwithstanding this, as the results from other
453 country-specific EQ-5D-5L valuation studies have shown a goodness-of-fit improvement and fewer insignificant
454 coefficients after excluding the flagged states (Ferreira *et al.*, 2019; Wong *et al.*, 2018), it will be valuable to analyse
455 the role of the Feedback Module on modelling SWEMWBS states with a larger sample size. Furthermore, some
456 potentially strategic DCE responses were not deleted because of insufficient evidence to prove the insincerity of
457 these answers. The presence of an interviewer was important to ensure the quality of these responses as the validity
458 of these responses could be judged by the degree of participants' engagement behaviour (e.g. the speed or mood of
459 completion).

460 Apart from separately modelling the C-TTO and DCE responses, this study also explored the IVWHM for
461 modelling both C-TTO and DCE responses (Model 3). We are not aware of the application of this approach within
462 previous health state preference elicitation studies. As this method attached more weight to parameters with lower
463 SEs, the correlation between the IVWHM value set and the DCE_{anchoring} value set (0.992) was higher than that
464 between the IVWHM value set and the C-TTO value set (0.957). Model 3 contained the least number of potentially
465 logical inconsistent coefficients and the lowest SEs surrounding the coefficients of attribute levels. Comparing
466 across all modelling approaches, this IVWHM offered a balanced perspective in gathering the preference

467 information elicited from two different valuation techniques. Across the pre-specified assessment criteria applied,
468 the IVWHM value set was preferred.

469 Whilst this study focused on the application of C-TTO and DCE methods for the valuation of the SWEMWBS,
470 recently developed novel valuation approaches should be considered by future research in this area. The main
471 feature of these approaches is that a social value set can be conveniently derived from the personal utility function
472 of each participant without the need to apply modelling techniques. For example, Devlin *et al.* (2019) elicited
473 personal utility functions by directly asking participants in England the relative importance of the EQ-5D
474 dimensions, levels and their associated interactions. Swing weighting was applied to allow participants to rate the
475 importance of dimensions based on the improvement from the worst level to the best level. According to the results
476 from the dimension and level rating tasks, they were given different pairs of choice tasks to decide the preferred
477 option between two states without duration. Questions comparing death to living in a health state for 10 years were
478 asked to derive the location of dead. Sullivan *et al.* (2020) asked participants adaptive DCE questions, selecting
479 between living in two health states with two EQ-5D-5L dimensions each within a given life span, followed by
480 questions to identify the diving threshold when moving from better-than-dead to worse-than-dead states. This form
481 of DCE questioning with fewer attributes within choice-pair comparisons could offer advantages for the valuation
482 of the SWEMWBS. The problem of overwhelming information caused by comparing seven-to-seven attributes with
483 different combinations of levels could be mitigated (Yiu *et al.*, 2022).

484 Furthermore, it was important to note one of the findings identified by both this study and the previous qualitative
485 study investigating the use of C-TTO and DCE for the valuation of the SWEMWBS (Yiu *et al.*, 2022), but not
486 documented in the previous health preference research. This relates to whether the state reflecting full health or
487 well-being should be valued highest. This appears logical for health utilities, where measures such as the EQ-5D-
488 5L capture impairment to physical or mental health dimensions, so the absence of such impairment should be
489 preferred. Therefore, the appropriateness of censoring responses with C-TTO value greater than one to an upper
490 limit of 1 within the HTM, though debatable, has a clear rationale, and an EQ-5D-5L quality assurance article
491 mentioned the impossibility of having TTO values greater than one (Alava *et al.*, 2020). However, the relative

492 desirability of full mental well-being was questioned by several participants during development of the SWEMWBS
493 value set (i.e. one out of 14 participants in the previous qualitative study and three participants out of 225
494 participants in this study) when completing C-TTO tasks. The reasons for not preferring “all of the time” for all
495 SWEMWBS items (e.g. preferring “often” for some items) were that the maximal well-being state was interpreted
496 as reflecting a lack of challenging life experience, which was a crucial element of an exciting and balanced life.
497 Respondents who expressed this view saw it as unhealthy to not have ups and downs in an individual’s mental life.
498 This argument would suggest that, unlike negatively worded HRQoL measures, it might be rational for some
499 participants to not assign the highest value to full mental well-being as individuals did not necessary think “always”
500 superior to “often”.

501 This issue had implications for the conduct and analysis of the C-TTO task. Three participants did not prefer full
502 mental well-being for some C-TTO tasks and they preferred life B (i.e. a state lower than full MWB) instead at the
503 beginning of the tasks. Their justification for not preferring full mental well-being was roughly similar to the reasons
504 raised by the participant in the qualitative phase, as described above. While it was possible to have a C-TTO value
505 greater than one for the valuation of SWEMWBS, only a small proportion of responses (17 responses, occupying
506 only 0.76% of the total responses) exhibited this non-monotonic preference (Supplementary Information 11.4). It
507 was therefore considered acceptable to censor those responses with C-TTO values greater than one to 1 when
508 modelling the C-TTO data in this study. However, future research could explore allowing C-TTO values greater
509 than 1 for some non-maximal states (which is equivalent to allowing full mental well-being to not be the highest
510 valued state). Also, a QALY unit and a MWALY unit have different relationships to common experience. One
511 QALY (e.g. living in full health for one year) is usually seen as a state that might be experienced by anyone who
512 does not have an illness (i.e. a normal unimpaired life). It is therefore a commonly experienced and plausibly the
513 most desirable state. Maximal mental well-being, however, is defined as experiencing the highest intensity of well-
514 being across all dimensions at all times. This is unlikely to reflect anyone’s lived experience, and it is less obvious
515 a priori that this would be optimal. It is therefore plausible for a close to highest mental well-being state to be more
516 preferable than the highest theoretically possible mental well-being state, at least to some respondents.

517 The limitation of this study is its sample size. Even though men and participants with low education levels were
518 underrepresented in the study sample, effort was exerted to purposely diversify the sample by including participants
519 with diverse sociodemographic characteristics. The final sample included participants that varied by gender, age,
520 education level, ethnicity, geographic region of residence and mental well-being status as described by the
521 SWEMWBS score. The lack of a financial incentive that could be offered to all participants and the inability to
522 conduct face-to-face interviews caused by social distancing measures discouraged potential participation. Several
523 insignificant coefficients for attribute levels might be attributable to limited statistical power of the C-TTO and
524 DCE models. Despite the smaller than ideal sample size, the coefficients generated by the HTM, CLM and IVWHM
525 were generally coherent and logically consistent with reasonably small SEs. Also, the results of the weighted
526 regressions with representative demographic characteristics of the general population were similar and consistent
527 to the unweighted regressions, suggesting that the sampling bias of having some non-representative demographic
528 characteristics in this study was minimal in affecting the ability of our modelling results to reflect public mental
529 well-being preferences in the UK. Even though the C-TTO model was shown to be relatively underpowered (i.e. it
530 generated the highest number of potentially logical inconsistent coefficients and statistically inconsistent
531 coefficients amongst models), the recommended value set generated by the IVWHM was similar to the value set
532 generated by the DCE model, which was sufficiently powered in this study based on the sample size estimation for
533 the DCE experimental design. Decision makers should interpret this preliminary value set for utility generation in
534 applied economic evaluations with caution given the current limitations. It will be beneficial for future research to
535 apply the methods developed in this study to derive a definitive value set from a more representative and larger
536 sample in the UK. Moreover, the decision around the types of covariates included in the modelling analysis was
537 informed by the results of the previous qualitative study (Yiu *et al.*, 2022). Future research could model the
538 interactions between the attribute levels and other demographic variables (e.g. income level). Furthermore, given
539 that the participants in the qualitative think-aloud study found that it was still manageable to compare alternative
540 combinations of levels for seven SWEMWBS attributes, we did not design overlapping attributes and levels for the
541 DCE or C-TTO tasks. Nonetheless, future research should continue to investigate the effect of setting overlapping
542 attributes and levels on the validity of valuation responses. Finally, problems related to installation process, remote

543 control, and display delay were discovered for the Microsoft Teams platform. The performance of other online
544 platforms for interviewer-administered preference elicitation studies should be explored.

545 **5. Conclusion**

546 This article documents the first attempt to derive preferences for SWEMWBS states. The IVWHM offered a
547 statistical way of combining both C-TTO and DCE data optimally. The IVWHM-derived value set can provide
548 useful indicative information to decision makers when comparing estimates of incremental cost per MWALY
549 gained generated by interventions, and to assist the identification of cost-effective interventions targeted at
550 improving mental well-being.

551

552 **6. References**

553 Alava, M. H., Pudney, S. & Wailoo, A. (2020) The EQ-5D-5L value set for England: findings of a quality assurance
554 program. *Value in Health*, 23 (5): 642-648.

555

556 Anthony, R., Moore, G., Page, N., Hewitt, G., Murphy, S. & Melendez-Torres, G. (2022) Measurement invariance
557 of the short Warwick-Edinburgh Mental Wellbeing Scale and latent mean differences (SWEMWBS) in young
558 people by current care status. *Quality of Life Research*, 31 (1): 205-213.

559

560 Bahrampour, M., Byrnes, J., Norman, R., Scuffham, P. A. & Downes, M. (2020) Discrete choice experiments to
561 generate utility values for multi-attribute utility instruments: a systematic review of methods. *The European Journal*
562 *of Health Economics*, 21 (7): 983-992.

563

564 Bass, M., Dawkin, M., Muncer, S., Vigurs, S. & Bostock, J. (2016) Validation of Warwick-Edinburgh mental well-
 565 being scale (WEMWBS) in a population of people using secondary care mental health services. *Journal of Mental*
 566 *Health*, 25 (4): 323-329.

567

568 Brazier, J. E. (2010) Is the EQ-5D fit for purpose in mental health? *British Journal of Psychiatry*, 197 (5): 348-349.

569

570 Brooks, R. & Group, E. (1996) EuroQol: the current state of play. *Health policy*, 37 (1): 53-72.

571

572 ChoiceMetrics (2018) *Ngene 1.2 USER MANUAL & REFERENCE GUIDE*. [online] Available from:
 573 <http://www.choice-metrics.com/NgeneManual120.pdf> (Accessed

574

575 Devlin, N. J., Shah, K. K., Mulhern, B. J., Pantiri, K. & van Hout, B. (2019) A new method for valuing health:
 576 directly eliciting personal utility functions. *The European Journal of Health Economics*, 20 (2): 257-270.

577

578 Diana Bardsley, Lucy Dean, Isla Dougall, Qingyang Feng, Linsay Gray, Malin Karikoski, Joe Rose, Caroline
 579 Stevens & Leyland, A. H. (2017) *The Scottish Health Survey: A National Statistics Publication for Scotland*.
 580 (Accessed 20 January 2019). Government, S.

581

582 Feng, Y., Devlin, N. J., Shah, K. K., Mulhern, B. & van Hout, B. (2018) New methods for modelling EQ-5D-5L
 583 value sets: An application to English data. *Health Economics*, 27 (1): 23-38.

584

585 Ferreira, P. L., Antunes, P., Ferreira, L. N., Pereira, L. N. & Ramos-Goñi, J. M. (2019) A hybrid modelling approach
586 for eliciting health state preferences: the Portuguese EQ-5D-5L value set. *Quality of Life Research*, 28 (12): 3163-
587 3175.

588

589 Greene, W. H. (2003) Chapter 21: Models for discrete choice. *Econometric Analysis, 5th ed. Upper Saddle River:*
590 *Prentice Hall*,

591

592 Haver, A., Akerjordet, K., Caputi, P., Furunes, T. & Magee, C. (2015) Measuring mental well-being: A validation
593 of the Short Warwick-Edinburgh Mental Well-Being Scale in Norwegian and Swedish. *Scandinavian Journal of*
594 *Public Health*, 43 (7): 721-727.

595

596 Helter, T. M., Kaltenboeck, A., Baumgartner, J., Mayrhofer, F., Heinze, G., Sönnichsen, A., Wancata, J. & Simon,
597 J. (2022) Does the relative importance of the OxCAP-MH's capability items differ according to mental ill-health
598 experience? *Health and Quality of Life Outcomes*, 20 (1): 99.

599

600 Jensen, C. E., Sørensen, S. S., Gudex, C., Jensen, M. B., Pedersen, K. M. & Ehlers, L. H. (2021) The Danish EQ-
601 5D-5L Value Set: A Hybrid Model Using cTTO and DCE Data. *Applied health economics and health policy*, 1-13.

602

603 John Brazier, Julie Ratcliffe, Joshua Salomon & Tsuchiya, A. (2017) *Measuring and Valuing Health Benefits for*
604 *Economic Evaluation*. Second edn. United Kingdom: Oxford University Press.

605

606 Johnson, R., Jenkinson, D., Stinton, C., Taylor-Phillips, S., Madan, J., Stewart-Brown, S. & Clarke, A. (2016)
607 Where's WALY? : A proof of concept study of the 'wellbeing adjusted life year' using secondary analysis of cross-
608 sectional survey data. *Health and Quality of Life Outcomes*, 14 9.

609

610 Koushede, V., Lasgaard, M., Hinrichsen, C., Meilstrup, C., Nielsen, L., Rayce, S. B., Torres-Sahli, M.,
611 Gudmundsdottir, D. G., Stewart-Brown, S. & Santini, Z. I. (2019) Measuring mental well-being in Denmark:
612 Validation of the original and short version of the Warwick-Edinburgh mental well-being scale (WEMWBS and
613 SWEMWBS) and cross-cultural comparison across four European settings. *Psychiatry Research*, 271 502-509.

614

615 Lancsar, E. & Louviere, J. (2008) Conducting discrete choice experiments to inform Healthcare decision making.
616 *Pharmacoeconomics*, 26 (8): 661-677.

617

618 Lee, C. H., Cook, S., Lee, J. S. & Han, B. (2016) Comparison of two meta-analysis methods: inverse-variance-
619 weighted average and weighted sum of Z-scores. *Genomics & informatics*, 14 (4): 173.

620

621 Louviere, J. J., Hensher, D. A., Swait, J. D. & Adamowicz, W. (2000) *Stated Choice Methods: Analysis and*
622 *Applications*. Cambridge University Press.

623

624 McFadden, D. (1973) Conditional logit analysis of qualitative choice behavior.

625

626 Ng Fat, L., Scholes, S., Boniface, S., Mindell, J. & Stewart-Brown, S. (2017) Evaluating and establishing national
627 norms for mental wellbeing using the short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS): findings
628 from the Health Survey for England. *Quality of Life Research*, 26 (5): 1129-1144.

629

630 Ng, S. S., Lo, A. W., Leung, T. K., Chan, F. S., Wong, A. T., Lam, R. W. & Tsang, D. K. (2014) Translation and
631 validation of the Chinese version of the short Warwick-Edinburgh Mental Well-being Scale for patients with mental
632 illness in Hong Kong. *East Asian archives of psychiatry*, 24 (1): 3-9.

633

634 Norman, R., Viney, R., Brazier, J., Burgess, L., Cronin, P., King, M., Ratcliffe, J. & Street, D. (2014) Valuing SF-
635 6D Health States Using a Discrete Choice Experiment. *Medical Decision Making*, 34 (6): 773-786.

636

637 Office for National Statistics *2011 Census: Key Statistics and Quick Statistics for Local Authorities in the United*
638 *Kingdom*. [online] Available from:
639 [https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/k](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/keystatisticsandquickstatisticsforlocalauthoritiesintheunitedkingdom/2013-10-11)
640 [eystatisticsandquickstatisticsforlocalauthoritiesintheunitedkingdom/2013-10-11](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/keystatisticsandquickstatisticsforlocalauthoritiesintheunitedkingdom/2013-10-11) (Accessed 12/10/2021).

641

642 Office for National Statistics *Population estimates for the UK, England and Wales, Scotland and Northern Ireland:*
643 *mid-2020*. [online] Available from:
644 [https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/a](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2020)
645 [nnualmidyearpopulationestimates/mid2020](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2020) (Accessed 12/10/2021).

646

647 Oppe, M. & Van Hout, B. (2017) The “power” of eliciting EQ-5D-5L values: the experimental design of the EQ-
648 VT. *EuroQol Working Paper Series*, 17003

649

650 Parkinson, J. (2007) Establishing a core set of national, sustainable mental health indicators for adults in Scotland:
651 Rationale paper. *Glasgow: NHS Health Scotland*,

652

653 Rogers, K. D., Dodds, C., Campbell, M. & Young, A. (2018) The validation of the Short Warwick-Edinburgh
654 Mental Well-Being Scale (SWEMWBS) with deaf British sign language users in the UK. *Health and Quality of*
655 *Life Outcomes*, 16 (1): 145.

656

657 Rowen, D., Brazier, J. & Van Hout, B. (2015) A Comparison of Methods for Converting DCE Values onto the Full
658 Health-Dead QALY Scale. *Medical Decision Making*, 35 (3): 328-340.

659

660 Saarni, S. I., Viertio, S., Perala, J., Koskinen, S., Lonnqvist, J. & Suvisaari, J. (2010) Quality of life of people with
661 schizophrenia, bipolar disorder and other psychotic disorders. *British Journal of Psychiatry*, 197 (5): 386-394.

662

663 Scottish Government (2018) *National indicator performance*. [online] Available from:
664 <https://nationalperformance.gov.scot/measuring-progress/national-indicator-performance> (Accessed 3 January).

665

666 Shah, K., Rand-Hendriksen, K., Ramos-Goni, J., Prause, A. & Stolk, E. (2014) Improving the quality of data
667 collected in EQ-5D-5L valuation studies: a summary of the EQ-VT research methodology programme.

668

669 Shah, K. K., Mulhern, B., Longworth, L. & Janssen, M. F. (2017) Views of the UK General Public on Important
670 Aspects of Health Not Captured by EQ-5D. *Patient-Patient Centered Outcomes Research*, 10 (6): 701-709.

671

672 Shah, N., Cader, M., Andrews, B., McCabe, R. & Stewart-Brown, S. L. (2021) Short Warwick-Edinburgh Mental
673 Well-being Scale (SWEMWBS): performance in a clinical sample in relation to PHQ-9 and GAD-7. *Health and*
674 *Quality of Life Outcomes*, 19 (1): 1-9.

675

676 Shah, N., Cader, M., Andrews, W. P., Wijesekera, D. & Stewart-Brown, S. L. (2018) Responsiveness of the short
677 Warwick Edinburgh mental well-being scale (SWEMWBS): evaluation a clinical sample. *Health and Quality of*
678 *Life Outcomes*, 16 (1): 1-7.

679

680 Stolk, E., Ludwig, K., Rand, K., van Hout, B. & Ramos-Goni, J. M. (2019) Overview, Update, and Lessons Learned
681 From the International EQ-5D-5L Valuation Work: Version 2 of the EQ-5D-5L Valuation Protocol. *Value in*
682 *Health*, 22 (1): 23-30.

683

684 Stolk, E. A., Oppe, M., Scalone, L. & Krabbe, P. F. M. (2010) Discrete Choice Modeling for the Quantification of
685 Health States: The Case of the EQ-5D. *Value in Health*, 13 (8): 1005-1013.

686

687 Sullivan, T., Hansen, P., Ombler, F., Derrett, S. & Devlin, N. (2020) A new tool for creating personal and social
688 EQ-5D-5L value sets, including valuing ‘dead’. *Social Science & Medicine*, 246 112707.

689

690 Vaingankar, J. A., Abdin, E., Chong, S. A., Sambasivam, R., Seow, E., Jeyagurunathan, A., Picco, L., Stewart-
691 Brown, S. & Subramaniam, M. (2017) Psychometric properties of the short Warwick Edinburgh mental well-being
692 scale (SWEMWBS) in service users with schizophrenia, depression and anxiety spectrum disorders. *Health and*
693 *Quality of Life Outcomes*, 15 (1): 153.

694

695 Wong, E. L. Y., Ramos-Goni, J. M., Cheung, A. W. L., Wong, A. Y. K. & Rivero-Arias, O. (2018) Assessing the
696 Use of a Feedback Module to Model EQ-5D-5L Health States Values in Hong Kong. *Patient-Patient Centered*
697 *Outcomes Research*, 11 (2): 235-247.

698

699 Yiu, H. H. E., Al-Janabi, H., Stewart-Brown, S., Petrou, S. & Madan, J. (2022) The use of composite time trade-
700 off and discrete choice experiment methods for the valuation of the Short Warwick-Edinburgh Mental Well-being
701 Scale (SWEMWBS): a think-aloud study. *Quality of Life Research*, 1-13.

702

703