



City Research Online

City, University of London Institutional Repository

Citation: Flood, C. ORCID: 0000-0001-5170-7792, Barlow, S. ORCID: 0000-0002-2737-8287, Simpson, A. ORCID: 0000-0003-3286-9846, Burls, A. ORCID: 0000-0001-9540-622X, Price, A., Cartwright, M. ORCID: 0000-0002-3404-5659, Brini, S. and Service User and Carer Group Advising on Research (SUGAR), (2018). What utility scores do mental health service users, healthcare professionals and members of the general public attribute to different health states? A co-produced mixed methods online survey. PLOS ONE,

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <http://openaccess.city.ac.uk/20772/>

Link to published version:

Copyright and reuse: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

City Research Online:

<http://openaccess.city.ac.uk/>

publications@city.ac.uk

1 **What utility scores do mental health service users,**
2 **healthcare professionals and members of the general**
3 **public attribute to different health states? A co-produced**
4 **mixed methods online survey**

5
6 (Short Title: Service user involvement in scoring utility measures)

7
8 Chris Flood^{1¶*}, Sally Barlow^{1¶}, Alan Simpson¹, Amanda Burls², Amy Price³, Martin
9 Cartwright², Stefano Brini². Service User and Carer Group Advising on Research (SUGAR)
10 members^{1^}

11
12 ¹ Centre for Mental Health Research, School of Health Sciences, University of London,
13 London and East London NHS Foundation Trust, London, United Kingdom

14 ² Centre for Health Services Research, School of Health Sciences, University of London,
15 London, United Kingdom

16 ³ Department of Continuing Education, University of Oxford, Oxford, United Kingdom

17
18 * Corresponding author:

19 Email: c.flood@city.ac.uk (CF)

20 ¶ These authors contributed equally to this work

21

22 ^ Membership of the Service User and Carer Group Advising on Research (SUGAR) is
23 provided in the Acknowledgements.

24

25 **Abstract**

26 **Background**

27 Utility scores are integral to health economics decision-making. Typically, utility scores have
28 not been scored or developed with mental health service users. The aims of this study were to
29 i) collaborate with service users to develop descriptions of five mental health states
30 (psychosis, depression, eating disorder, medication side effects and self-harm); ii) explore
31 feasibility and acceptability of using scenario-based health states in an e-survey; iii) evaluate
32 which utility measures (standard gamble (SG), time trade off (TTO) and rating scale (RS))
33 are preferred; and iv) determine how different participant groups discriminate between the
34 health scenarios and rank them.

35

36 **Design and methods**

37 This was a co-produced mixed methods cross-sectional online survey. Utility scores were
38 generated using the SG, TTO and RS methods; difficulty of the completing each method,
39 markers of acceptability and participants' preference were also assessed.

40

41 **Results**

42 A total of 119 participants (58%) fully completed the survey. For any given health state, SG
43 consistently generated higher utility scores compared to RS and for some health states higher

44 also than TTO (i.e. SG produces inflated utility scores relative to RS and TTO) . Results
45 suggest that different utility measures produce different evaluations of described health
46 states. The TTO was preferred by all participant groups over the SG. The three participant
47 groups scored four (of five) health scenarios comparably. Psychosis scored as the worst
48 health state to live with while medication side-effects were viewed more positively than other
49 scenarios (depression, eating disorders, self-harm) by all participant groups. However, there
50 was a difference in how the depression scenario was scored, with service users giving
51 depression a lower utility score compared to other groups.

52

53 **Conclusion**

54

55 Mental health state scenarios used to generate utility scores can be co-produced and are well
56 received by a broad range of participants. Utility valuations using SG, TTO and RS were
57 feasible for use with service users, carers, healthcare professionals and members of the
58 general public. Future studies of utility scores in psychiatry should aim to include mental
59 health service users as both co-investigators and respondents.

60

61 **Keyword**

62 Utility Measure, Time Trade-off, Standard Gamble, Service User Involvement

63 **Introduction**

64

65 Mental ill health is a key contributor to the burden of disease [1] costing an estimated £70-
66 £100 billion per year in the United Kingdom (UK), equivalent to 4.5% of gross domestic
67 product (GDP) [2]. Over half of this cost relates to reduced quality of life [3]. There is a need
68 to prioritise interventions that are cost-effective and target health states that service users
69 report have the greatest impact on their lives.

70

71 Traditionally, health economists and policy-makers use health utilities to estimate treatment
72 cost-effectiveness and inform prioritizing decision-making [4]. Utility is a weighted and
73 scaled method of quantifying a person's preference for an experienced or hypothetical health
74 state [5]. Utility scores are obtained by asking people to evaluate their preference for living in
75 particular health states (e.g., depression) or experiencing a health-related event (e.g.,
76 medication side effects). This evaluation may draw on a person's current or past experience,
77 or their imagining of what it would be like to live with the health state in the future [6].

78 Service users who live with mental illness/distress and receive mental health treatments are
79 well-placed to inform policymakers on the impact of mental ill health on their quality of life.

80

81 Utility scores can be elicited using various methodologies including time trade-off (TTO),
82 standard gamble (SG) and rating scales (RS) [7]. These methods ask people to consider
83 hypothetical health states and either trade an improvement in health for a reduction in time
84 alive (TTO) or a greater risk of death (SG), or to rate the health state on a scale (RS). Several
85 factors are important when choosing a utility method to use. These include their
86 psychometric properties such as: validity (does the method elicit a true preference for the
87 health state?); reliability (does the method elicit reproducible scores?); feasibility (is the

88 method practical for the target population and setting?); and acceptability for the target
89 population. In situations where there is relatively little experience with making health state
90 valuations it is desirable to employ several methods in parallel to determine which is the most
91 suitable [8]. However, when employed in parallel, different methods can yield different
92 utility scores raising legitimate concerns about how patients, commissioners and policy-
93 makers should use the evidence from different utility methods to guide their decisions [9, 10].

94
95 The SG is considered the ‘gold standard’ method because it includes an element of
96 uncertainty, thought to reflect real world uncertainty over decisions about health and
97 healthcare. The TTO and RS do not involve uncertainty but rather derive utility values which
98 may be transformed to utility scores [11]. Some studies report that the SG method generates
99 higher scores (indicating more positive evaluations, or less negative evaluations, of the health
100 state) than TTO and RS [12]. Similarly some have found that TTO is scored higher than RS
101 [7].

102
103 Whether societal preference (amongst the general population) or experience-informed
104 preference (amongst patients) should guide policy-making remains contested [4]. Gold et al.
105 (1996) propose that societal preferences should be used for macro-level decision-making and
106 patient preferences for meso-level (guideline development) decision-making [13]. Utility
107 scores derived from patients may differ from those of the general population or other
108 specialist groups such as healthcare professionals. For example, people experiencing multiple
109 health states give greater weight to mental health states than physical health states, compared
110 to the scoring of the general population [4, 14]. However, some authors have raised concerns
111 over the challenges of producing fair and balanced evaluations of health states for individuals
112 who have personally experienced the health state or symptoms described [15].

113

114 Utility scores are widely used for priority setting and resource allocation for physical health
115 states, but less frequently for mental health states [11, 16]. However, there is little evidence
116 of service users' inclusion in the development or scoring of valuations in mental health states
117 [11]. Emphasis has been placed on the cognitive challenge that scoring health utilities poses
118 and how some mental illnesses may limit comprehension of the task [17]. Despite these
119 concerns, empirical studies demonstrate it is feasible to derive health utility scores from
120 patients with severe and enduring mental illnesses such as schizophrenia [15, 17-19], bipolar
121 disorder [20], depression [11] and affective and alcohol related disorders [15]. These studies
122 have demonstrated that service users can discriminate by disease severity and medication side
123 effects [20]. It has been recognised that the questions and procedures used to generate utility
124 scores are abstract and challenging [21] and there have been recommendations that
125 methodology should be refined to accommodate patients' 'mental status' [17]. Another
126 significant concern around framing effect biases, can be reduced through the involvement of
127 mental health service users in developing the health state descriptions used to elicit utility
128 scores, [7] and is central to the research approach reported in this paper.

129

130 A comparative research design, with study materials co-produced with service users, may
131 reduce some of the limitations to help achieve more valid utility measures. Studies
132 comparing SG and TTO are usually exclusively quantitative. Our study also includes
133 qualitative elements to gain insight into the acceptability of the different utility methods and
134 explore factors influencing participants' values and preferences in health state valuations
135 [22]. We also examine whether health utilities can be measured remotely using an e-survey.

136

137 Study objectives were:

- 138 - To co-develop descriptions of mental health states from which utility scores could be
139 derived, and co-produce utility questions that are understandable to service users.
- 140 - To assess the feasibility and acceptability of using scenario-based health states to
141 measure health utility.
- 142 - To determine which utility measure is preferred and how participant groups
143 discriminate between the scenarios.
- 144 - To compare utility scores provided by service users, carers, healthcare professionals
145 and interested members of the general public.

146

147

148

149

150 **Materials and Methods**

151 **Research design**

152 The study used a cross-sectional online survey to collect quantitative and qualitative data.

153 **Population, sampling and data collection**

154 Mental health service users, carers, healthcare professionals and interested members of the public

155 were invited to take part in the survey. Service users and carers were recruited via a link on the

156 national Rethink Mental Illness charity website (www.rethink.org). The survey was promoted by

157 snowball emails and social networking sites Twitter and Facebook. Participants were self-selecting

158 and indicated which participant group they identified with. The electronic survey was open for

159 recruitment from March 2015 to July 2015. The study was conducted in the UK but did not preclude

160 participation from other countries.

161

162 **Ethical Approval**

163 Ethical approval was granted by the School of Health Sciences' Research Ethics Committee
164 City, University of London.

165

166 **Instrument Design**

167 The survey was designed collaboratively with members of the Service User and Carer Group
168 Advising on Research (SUGAR) [23] and research academics at City, University of London.

169 The SUGAR group has 13 service users with lived experience of mental illness and three
170 carers and meets monthly to advise on research projects within the Centre for Mental Health

171 Research and East London NHS Foundation Trust. The original study design was presented

172 at the January 2013 SUGAR group meeting. Members were invited to become involved in

173 the study, to ask questions about the research and discuss how the research study could

174 proceed. Members of the SUGAR group agreed to work collaboratively on the study and

175 contribute to the design of instruments. Instrument design occurred in two stages: 1)

176 developing the mental health state scenarios; and, 2) designing the survey questions.

177

178 **Stage 1: Development of the Mental Health State Scenarios**

179 The SUGAR group helped write several short fictional scenarios describing the presentation
180 and experience of specific mental health conditions. Members worked in groups of two or

181 three. We offered guidance to the group by prompting members with questions such as 'how
182 would someone describe living with that condition?' 'What would impact on their

183 condition?' Once complete, each scenario was presented to the wider group for feedback. The

184 scenarios went through several iterations through group discussions over three months. An

185 example hypothetical health state and its description is given in box 1.

186

187

Box 1: Psychosis health state scenario

188

189

190

'Joseph lives alone and is scared that people are out to kill him and says that these people are going to bomb his house. His neighbours also want him out because of what they see as strange behaviour on his part, his general oddity and the fact that he talks to himself. Joseph hears voices which reinforce his fears.'

191

192

193

194

195

196

A total of ten scenarios, focusing on different mental health states, were developed. The final survey used five scenarios chosen by the SUGAR group. The example in box 1 focusing on psychosis, and another four scenarios on medication side-effects, self-harm, eating disorders and severe depression were used (see [S15](#) Table).

197

Stage 2: Development of the survey questions

198

199

200

201

Survey questions were designed using SG and TTO methods for scoring the five health states. The survey questions were designed by the authors and reviewed by the SUGAR group to ensure that they were comprehensible.

202

203

204

205

206

207

208

209

210

The survey was housed on SmartSurvey (www.smartsurvey.co.uk) and included 29 questions (some with several parts). We estimated it would take 20-25 minutes to complete. The first part asked respondents to read the information sheet and consent to the study. Participants could withdraw from the study at any stage by simply clicking out of the survey. For descriptive purposes socio-demographic information (i.e., age, gender, ethnicity, level of education and marital status) was requested and is summarised in table 1. The main body of the survey included the five scenarios with questions linked to each of the scenarios to assess health utilities using RS, SG and TTO methods. After completing the three utility measures for all five scenarios, respondents were asked questions about the acceptability of the

211 measures. Ten-point Likert rating scales assessed the perceived difficulty of each method.
212 Preference for each method was reported alongside free text response boxes so that
213 participants could expand on their responses. A final free-text response box at the end of the
214 survey allowed for feedback on anything that may have affected their response to the
215 questions. The Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [24] was
216 used to inform the development of the survey.

217

218 **Utility Measures**

219 All the utility methods generate a score from 0-1 (0: *worst possible health state* – 1: *best*
220 *possible health state*). The methods of eliciting utility scores for each measure are described
221 below:

222

223 **Rating Scale questions**

224 For each scenario participants were asked to score the health state from 0-10 with lower
225 values representing more negative appraisals of the health state (0: *worst imaginable health*
226 *state* – 10: *the best possible health state*). In order to obtain a RS utility measure the
227 responses given by the participant was divided by 10 to produce values between the ranges of
228 0-1.

229

230 **Time Trade-Off questions**

231 For each scenario respondents were asked to imagine making a choice between spending the
232 next ten years of life in the health state described (e.g., psychosis), or 'trading' some years of
233 life to be completely free of symptoms for the rest of their life. They were then asked to
234 indicate the maximum number of years of their life they would be willing to trade to have
235 complete wellness. To help with comprehension the SUGAR group suggested that some

236 people might understand the term 'trade' as swapping, surrendering or sacrificing, this was
237 incorporated into the description.

238

239 The choice of how many years to trade was offered incrementally, one year at a time (to a
240 maximum of 10). A choice was required for each year to identify the point of indifference
241 which was reached when the participant could no longer choose. The utility for the health
242 state was calculated from the proportion of years traded at the point of indifference. For
243 example, if someone trades 4 out of a possible 10 years of life to achieve full health, then the
244 utility they ascribe to the health state would be 0.6 (Utility = $1 - (\text{years traded at the point of}$
245 $\text{indifference}/\text{total possible years to trade})$).

246

247 **Standard Gamble (SG) questions**

248 For each scenario respondents could choose to remain in the health state (e.g., psychosis) for
249 the rest of their lives or take a gamble in which there was a specified risk of dying but, if they
250 did not die, they would be fully healthy. As part of this process respondents were asked to
251 score the maximum risk of death they would take in exchange for guaranteed full health until
252 a point of indifference. For example, if respondents find it hard to choose whether or not they
253 would risk a 10% chance of death for a 90% chance of full health, then their utility for that
254 state is 0.9. If they are indifferent when there is a 90% chance of death and 10% chance of
255 full health, then their utility for that health state is 0.1.

256

257 **Analysis**

258 **Quantitative analysis**

259 Utility scores were calculated in Microsoft Excel, imported into SPSS version 21 [25] and
260 checked and cleaned by two researchers. Descriptive statistics (e.g., means + standard
261 deviations; frequencies; percentages) were used to summarise the sample characteristics and
262 the outcome measures (utility scores) using three different methods (RS, TTO, SG) for five
263 health states, across three participant groups (SU, HCP, MoP). These comparisons enable an
264 evaluation of discriminatory power. Discriminatory power is a function of three factors: the
265 description of the health state, the utility method, and the evaluative abilities of participants.
266 Discriminatory power is present when health states that would be expected to be scored
267 differently are scored differently. Observing discriminatory power therefore implies that no
268 influential biases are present (e.g., floor or ceiling effects, central tendency bias). Conversely,
269 a lack of discriminatory power across all participant groups raises questions about the health
270 state descriptions and/or the utility method, whereas a lack of discriminatory power in only
271 some participant groups suggests a lack of evaluative abilities in those groups.

272

273 **Understanding, acceptability and preference**

274 Understanding, acceptability and preference for the three utility scoring methods across the
275 different participant groups was assessed using (a) the proportion of successfully completed
276 surveys, (b) a perceived difficulty Likert scale, (c) the reported preference for the utility
277 measures, and (d) the number of zero traders and maximal traders.

278

279 **Statistical analysis**

280 Inferential statistics were based on three groups (n=105): service users (SUs; N=46),
281 healthcare professionals (HCPs; N=28) and interested members of the public (MoPs; N=31)
282 because there were insufficient participants in other groups (Carers; N=6, Service users and
283 healthcare professionals; N=5). A two-way mixed analysis of variance (ANOVA) was

284 conducted to determine whether there were significant differences in the way the three groups
285 (SUs, HCPs, MoPs) scored the five health states using the three types of utility measure (RS,
286 TTO or SG). Due to multiple testing, the level of significance (α -level) was reduced to 0.01.
287 Tukey post-hoc tests were conducted to ascertain differences in scoring for the different
288 scenarios, utility measures and participant groups.

289

290 One-way ANOVAs were conducted to explore differences in the perceived difficulty of each
291 utility measure and the percentage preference scores, across groups.

292

293 **Analysis of Zero-traders and Maximal Traders**

294 Zero-traders, respondents who did not trade any years of life for improved health (TTO) or
295 gamble at any % risk of death (SG), and maximal-traders, respondents who traded the
296 maximum amount of time (10 years) or accepted the maximum amount of risk to live in
297 perfect health, were identified.

298

299 **Qualitative Data**

300 The free text boxes enabled participants to provide qualitative information about factors that
301 may have influenced their responses and their preference of utility measure. A basic thematic
302 analysis was undertaken [26] line-by-line using constant comparisons. Identified themes were
303 independently checked by two researchers and disagreements resolved by a third reviewer.

304

305

306 **Results**

307 **Sample**

308 During the four month recruitment period 204 people accessed the survey: 85 were partially
309 completed and 119 (58%) fully completed. The mean time to complete the survey was 14
310 mins (range from 4 mins to 120 mins), with 75% of respondents completing the survey
311 within 9 and 57 minutes. Eight participants returned to the survey and time of completion
312 could not be obtained. Participant characteristics are given in table 1. Of the 119 complete
313 responders, 46 identified as service users; 6 were carers; 31 were interested members of the
314 public; and 28 were mental health professionals. Five respondents described themselves as
315 both a service user and a healthcare professional. A further 3 participants that selected
316 multiple identities were excluded, leaving 116 in the descriptive data. Participants were
317 between 17 and 72 years old. More females completed the survey than males, ranging from
318 68%-100% across participant groups. The majority of the respondents were based in England
319 and a large percentage (73%) identified themselves as English or “other British”. A high
320 proportion had University-level education, with 22% reporting having a degree (e.g., BA,
321 BSc) and 37% a higher degree (e.g., MSc, PhD).

322
|

324 **Table 1. Participant socio-demographic data**

	Service Users (n=46)	Carers (n=6)	Interested member of the Public (n=31)	Healthcare Professionals (n=28)	Service Users & Health Care Professionals (n=5)
Age in years: mean (s.d.)	32 (12)	49 (17)	36 (11)	39 (12)	40 (18)
Range	17-62	19-72	18-59	21-57	29 -67
Gender:					
Female	34 (74)	5 (83)	29 (93)	19 (68)	100 (5)
Male	12 (26)	1 (17)	2 (7)	9 (32)	
Country					
England	42 (91)	5 (83)	30 (97)	25 (89)	5 (100)
Wales	2 (4)				
Scotland	1 (2)	1 (17)		1 (4)	
Other country	1 (2)		1 (3)	2 (7)	
Ethnicity:					
English	30 (65)	4 (67)	18 (58)	17 (61)	3 (60)
Other British	7 (15)	1 (17)	4 (13)	2 (7)	1 (20)
Other White			2 (6)	1 (4)	
Asian			4 (13)	3 (14)	
Irish	2 (4)			1 (4)	
African			1 (3)	3 (11)	
Black/British	1 (2)	1 (17)			
Black/Caribbean			1 (1)		
Other ethnic group	3 (6)		1 (3)		
Relationship Status:					
Never married/formed a civil partnership	27 (59)		11 (35)	10 (36)	3 (60)
Married/in civil partnership	9 (20)	5 (81)	9 (29)	7 (25)	
Cohabiting	7 (15)	1 (17)	3 (10)	9 (32)	2 (40)
Divorced/Separated	2 (4)		5 (16)	2 (7)	
Widowed	1 (2)		3 (10)		
Work Status:					
in paid employment	25 (48)	3 (50)	17 (55)	22 (79)	3 (60)
temporarily off sick	4 (8)				
Unemployed	2 (4)	1 (17)	3 (10)		
Retired		1 (17)	2 (6)		
looking after the family, home or dependents			2 (6)		
Unable to work because of Long term disability or ill health	6 (13)				
In full time education or training	9 (20)	1 (17)	6 (19)	6 (21)	2 (40)
Other	3 (6)		1 (3)		
Qualification:					
Higher degree	17 (37)	2 (33)	14 (45)	17 (61)	1 (20)
Degree/degree level	10 (22)		12 (39)	10 (36)	
Other higher education below degree	4 (9)		1 (3)	1 (4)	3 (60)
A-levels/similar	9 (17)	1 (17)	2 (6)		1 (20)
GCSE/O-level/similar	5 (11)	2 (33)	2 (6)		

Trade Apprenticeships	1 (2)	
No Qualifications	1 (2)	1 (17)

326 *All data presented as N and (%) unless stated otherwise. Missing data for Age, N=1, Service user and Health*

327 *care professional*

328

329

330

331

332 **Non-completion of the survey**

333 The surveys that were started but not completed (N=85) were not included in the analysis
334 however we provide some further detail here. Seventy nine completed the first stage of the
335 survey allowing us to view the socio-demographics. The majority left the survey after
336 completing the first scenario questions. The demographics of the participants completing the
337 survey were similar to those who did not. The average age of non-completers was 36, 45/79
338 were female, 44/79 had a higher degree and 24/79 were service users.

339

340 **Comparative utility scores**

341 Utility scores ranged from zero to one. Comparative mean scores for the utility measures and
342 participant groups are provided in table 2. Similar patterns of scoring were observed across
343 participant groups and the SG consistently scored higher (indicating a better health state) than
344 the RS in all five scenarios, and for some health states more than TTO.

345

346 There were no significant interactions between utility measure and the participant group for
347 any scenario. There was a substantial main effect of utility measure on utility score in all
348 scenarios, suggesting that different utility measures produce different scores. Table 2
349 summarises the descriptive and inferential statistics. There were no significant differences in
350 how participant groups scored four scenarios (psychosis, side-effects, self-harm and eating
351 disorders). There was a significant main effect of participant group in the depression
352 scenario, $F(2, 102) = 4.80$, $p = 0.01$, partial eta squared = 0.086, suggesting that there was a
353 difference in the way that service users, healthcare professionals and interested members of
354 the public scored this scenario. Tukey post-hoc tests suggested that service users gave

355 depression a lower utility score (RS: 0.30; TTO: 0.31; SG: 0.50) (perceived it as worse to live
356 with) than healthcare professionals (RS: 0.33; TTO: 0.49; SG: 0.70), $p = 0.036$ and interested
357 members of the public (RS: 0.36; TTO: 0.48; SG: 0.64), $p = 0.025$ (post hoc tests non-
358 significant at reduced α level < 0.01).

359 **Table 2. Comparative utility scores between utility measures used and respondents**

Utility Measures by Scenario (Mean ± SD)	Participant Groups		
	Mental Health Service User (n=46)	Interested member of the public (n= 31)	Healthcare Professional (n=28)
Psychosis			
RS	0.30 (0.17)	0.26 (0.22)	0.26 (0.16)
TTO	0.24 (0.33)	0.28 (0.29)	0.34 (0.33)
SG	0.50 (0.37)	0.50 (0.26)	0.52 (0.34)
ANOVA:			
by Group	F (2, 102) = 0.189, p = 0.83		
by Utility	F (1.95, 204) = 24.65, p<0.0001		
Medication Side-effects			
RS	0.47 (0.18)	0.53 (0.19)	0.53 (0.15)
TTO	0.56 (0.36)	0.65 (0.31)	0.68 (0.33)
SG	0.68 (0.37)	0.74 (0.33)	0.78 (0.33)
ANOVA:			
by Group	F (2, 102) = 1.67, p = 0.19		
by Utility	F (2, 204) = 22.01, p<0.0001		
Self-Harm			
RS	0.38 (0.20)	0.36 (0.19)	0.39 (0.20)
TTO	0.47 (0.39)	0.53 (0.30)	0.58 (0.31)
SG	0.61 (0.36)	0.67 (0.32)	0.75 (0.29)
ANOVA:			
by Group	F (2, 102) = 1.32, p = 0.27		
by Utility	F (2, 204) = 41.05, p<0.0001		
Eating disorders			
RS	0.35 (0.22)	0.35 (0.20)	0.41 (0.19)
TTO	0.42 (0.34)	0.56 (0.33)	0.63 (0.31)
SG	0.60 (0.34)	0.69 (0.31)	0.75 (0.28)
ANOVA:			

	by Group	F (2, 102) = 3.65, p = 0.03		
	by Utility	F (2, 204) = 46.85, p<0.0001		
Depression				
	RS	0.30 (0.21)	0.36 (0.23)	0.33 (0.18)
	TTO	0.31 (0.35)	0.48 (0.30)	0.49 (0.34)
	SG	0.50 (0.36)	0.64 (0.29)	0.70 (0.25)
ANOVA:				
	by Group	F (2, 102) = 4.80, p = 0.01		
	by Utility	F (2, 204) = 35.67, p<0.0001		

360

361

362

363 **Ranking the scenarios**

364 There was considerable consistency in how the scenarios were ranked and the type of utility
365 measure used (see Table 3). Similarly, there was consistency across the participant groups'
366 mean ranking of health states.

367

368 **Table 3 – Health State Scenarios Ranked according To Valuation by Method and Participant**

Mental Health Service Users (N= 46)			
Rank	Rating Scale	Time Trade-Off	Standard Gamble
1	Psychosis / Depression	Psychosis	Psychosis / Depression
2	Eating Disorders	Depression	Eating Disorders
3	Self-harm	Eating Disorders	Self-Harm
4	Medication side-effects	Self-Harm	Medication side-effects
5		Medication side-effects	
Interested members of the public (N=31)			
Rank	Rating Scale	Time Trade-Off	Standard Gamble
1	Psychosis	Psychosis	Psychosis
2	Eating Disorders	Depression	Depression
3	Self-Harm / Depression	Self-Harm	Self-Harm
4	Medication side-effects	Eating Disorders	Eating Disorders
5		Medication side-effects	Medication side-effects
Healthcare Professionals (N=28)			
Rank	Rating Scale	Time Trade-Off	Standard Gamble
1	Psychosis	Psychosis	Psychosis
2	Depression	Depression	Depression
3	Self-Harm	Self-Harm	Self-harm / Eating Disorders
4	Eating Disorders	Eating Disorders	Medication side-effects
5	Medication side-effects	Medication side-effects	

369 Rank: 1 = Worst Health State

370

371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391

Across all groups and utility measures, psychosis scored as the worst health state to live with while medication side-effects were viewed most positively. MoPs and HCPs scored depression as the second worst health state across all utility measures; SUs ranked depression equal to psychosis using the RS and SG. Eating disorders and self-harm were mid-ranked across all groups and utility measures.

Acceptability of the utility measures

Perceived difficulty in completing the questions.

Participants were asked to measure on a Likert scale how hard they thought it was to complete the questions. A score of zero referred to ‘not difficult at all’ and a score of 10 represented ‘very difficult’. A one-way ANOVA revealed no significant differences in the perceived difficulty of the utility methods between SUs (mean = 5.35 (SD = 2.87), MoP (6.32, 2.86) and HCPs (6.75, 2.44) ($F(2, 102) = 2.53, p = 0.085$).

Further qualitative detail about the perceived difficulty in completing the questions was derived from the free-text responses from 38 participants (12 SUs, 1 Carer (C), 14 HCPs, 7 MoP, and 4 people identifying with two or more of the population categories). These were collated into five core themes, a summary is provided in table 4 with illustrative quotes.

Moral and emotional reactions

Participants referred to how they reflected on their choices and spoke about emotional reactions to the questions and moral dilemmas that they felt when completing the valuations.

- *'frustrated as I couldn't explain my choices'* [C & HCP]
- *'The questions which raised isolation as a factor made me more likely to trade years'* [HCP]
- Another spoke about feeling *'despair'* [C]
- *'I felt guilty rating things as less important as it seemed like I was belittling the condition'* [P]
- *'questions difficult in a moral sense'* [HCP]

Relevance to own experience

Some participants argued that lived experience could be advantageous in answering the questions. Concerns were raised about difficulties in imagining what it would be like to live with some health state. This was acknowledged by SUs and HCPs.

- *'Could relate more to own experience so rated them worse, which makes my answers subjective, rather than objective'* [SU]
- *'Difficult to understand what those symptoms really feel like and be able to accurately make a judgment as to what you would do'* [HCP]
- *'As only one section was even vaguely relevant to something I had experienced, I did not feel competent to make an assumption of what it would be like to experience most of the states described'* [SU]
- *'finding it hard, to imagine being in the described situations'* [HCP]

Standard Gamble Confusing

Several people found the wording in the risk question difficult

- *'percentage questions confusing'* [HCP]
- *'finding the 'risk % section quite hard to understand'* [P]
- *'Risk % section quite hard to understand'* [SU]

Instructions unclear/ambiguous

Several responses were received around the wording and difficulty with interpreting what was expected when completing the valuations.

- *'I found the wording & the concept of the questions confusing'* [HCP]
- Another mentioned *'not really understanding your instructions. It felt very abstract.'* [P]
- *'Instructions were complex'* [P] *'Too complicated'* [P]
- *'Instructions were not clear'* [HCP]

Conceptually challenging/uncertainty over choices

Some respondents found the methods conceptually challenging and making valuations philosophically difficult. Some references were made to concerted efforts in thinking through the responses and making judgements.

- *'This is not a questionnaire that could easily be completed by a lay person who does not have research training'* [P]
- *'hard to be consistent across questions'* [P]
- *One respondent stated it was 'hard to make a judgement between trading the end years of your life with the likelihood of dying by suicide in the next ten years'* [SU]
- *'Life & death decisions are hard and not very realistic'* [SU]

395

396

397 **Utility Measure Preference**

398 The majority of participants (N= 72, 60%) found the TTO measure easier to complete than
399 the SG (N=47, 40%). Exploring differences across participant groups revealed that
400 preference for the TTO was held by SUs(63%, N=29), MoP(58%, N=18), HCPs(61%, N=17)
401 and 3 out of 5 (60%) people identifying as both SUs andHCPs.

402 There were 22 free-text responses about how the preferred choice was made. Ten were from
403 SUs, six fromMoP, four from HCPs, and one each from a carer and a SU who was also
404 aHCP. Nine respondents indicated that neither TTO nor SG was easier to complete, three
405 reported that SG was easier, and four thought that the TTO was easier and gave reasons for
406 these.

407 Participants who preferred the TTO measure (N=4) provided responses that fell into two
408 main categories:

409

410 - **Lack of clarity (of SG):** TTO was easier to understand than the SG because “[SG
411 was] *confusingly worded*” and “*a bit too arbitrary*”.

412

413 - **Personal meaning:** “Would rather have quality of life over duration” [service user
414 and healthcare professional]. A service user expressed that the TTO was easier to
415 relate to for them “*Because I know how much time in my life has been lost being ill*”.

416

417 Participants who preferred the SG (N=3) provided responses that could be grouped into two
418 key areas of concern.

419

420 - **Uncertainty:** not knowing the length of life years that they had left : ‘[*finding*]

421 *Balance between trading time for wellness is difficult to assess given none of us know*

422 *how long we will live. Also is effected by age. The percentage risk is more immediate'*
423 [SU].

424

425 - **Complexity:** A service user thought that the wording in the TTO was more difficult
426 *"I couldn't figure out whether I would spend 10 years unwell and then be okay for the*
427 *rest of my life or 10 years and then die straight away."* A mental health professional
428 also stated that the SG was easier because they were *"short questions and easy to*
429 *select the percentage'*.

430

431 **Analysis of Zero-traders and Maximal-traders**

432 Zero-traders are participants who want the maximum length of life at whatever cost to quality
433 of life. Maximal-traders want the maximum quality of life at whatever expense to length of
434 life. There were zero-traders and maximal-traders in both of the utility scoring methods,
435 these will be presented in turn.

436

437 **Time Trade-Off**

438 Five participants were zero-traders across all scenarios using the TTO method (2 service user,
439 2 healthcare professionals and 1 member of the public). Overall there were 70 incidents of
440 zero trading (12% of all responses) across the scenarios and participants. The highest
441 incidence of zero traders was for the medication side-effects scenario with 24 participants
442 (20%) choosing not to trade years. This contrasts with only 6 participants (5%) choosing not
443 to trade years in the psychosis scenario.

444 Seven participants who were maximal traders across all scenarios using the TTO method (4
445 service users, 1 carer, 2 healthcare professionals and 1 member of the public). There were
446 152 incidents of maximal trading (26% of all responses) across the scenarios. The highest

447 incidence of maximal traders was in the psychosis scenario with 58 participants (49%)
448 choosing to trade the maximum number of years (10 years) to live in a better health state.
449 The lowest incidence of maximal trading was for medication side effects with 16 people
450 (13%) choosing to trade 10 years to live without side effects.

451

452 **Standard Gamble**

453 Three participants were zero-traders across all scenarios using the SG method and all were
454 service users. Overall there were 69 incidents of zero trading (12% of all responses) across
455 the scenarios and participants. The highest incidence of zero traders was for the medication
456 side-effects scenario with 26 participants (22%) choosing not to accept any % risk of death
457 for a better health state. This contrasts with only 6 participants (5%) choosing not to risk
458 death in the psychosis scenario.

459 Six participants were maximal traders across the scenarios using the SG method (3 service
460 users, 1 carer and 2 members of the public). Overall there were 62 incidents of maximal
461 trading (10%) across the scenarios. The psychosis scenario had the highest incidence of
462 maximal traders (17/119 (14%)), while medication side effects had the lowest incidence of
463 maximal trading (11/119 (9%)).

464 **Did anything else affect participants' responses?**

465

466 We received 66 participant responses to the open-ended question asking if anything had
467 affected their responses. The majority of the responses (N= 46, 70%) related to personal
468 experiences of mental illness and identifying with the person in the scenario. Twenty four of
469 the responses about personal experience were from service users, with seven from HCPs, six
470 from MoP, three from carers, and five responses from people identifying with two or more of
471 the participant type categories.

472

473 *“Probably those ones that I can 'feel' the pain of relative to those ones that I have to*
474 *imagine - I'm probably more willing to trade years on things that I can remember*
475 *feeling.” [SU]*

476

477 *“Partner suffers from psychosis and I have seen this suffering straight on” [C]*

478

479 *“I have experienced severe depression myself & have also worked with people with*
480 *the rest of the diagnoses discussed.” [HCP]*

481

482 In contrast, others reflected on their lack of personal lived experience of mental illness and
483 how that brought challenges in completing valuations on the health states.

484

485 *“Not having first-hand experience and therefore having to rely on impressions of my*
486 *personality to consider what my actions might be” [MoP]*

487

488 Several respondents referred to their emotional state at that moment (N=7), mentioning
489 feelings of sadness, tiredness and social isolation. Whilst two people reflected on the
490 complexity of the scenarios and others on challenges with moral decisions and two service
491 user participants referred to negative images or stigma of mental health conditions.

492 **Discussion**

493 In this study, we sought to collaborate with service users to co-produce descriptions of mental
494 health states from which to generate utility scores and frame utility questions so that they are

495 comprehensible to service users. Another aim was to determine the feasibility of using
496 different utility methods via an online questionnaire. We compared utility scores provided by
497 service users, healthcare professionals, members of the public, and carers (descriptively).
498 The acceptability of the co-produced health states and the different utility methods to
499 determine health utility was also examined.

500

501 The results indicated that:

- 502 1. Mental health state scenarios used to generate utility scores can be co-produced and
503 are well received by a broad range of participants using an online survey.
- 504 2. Standard techniques used to elicit utility valuations (SG; TTO and RS) were feasible
505 for use with service users, carers, healthcare professionals and members of the general
506 public.
- 507 3. Similar trends were seen in utility scores elicited by the different utility methods
508 across all participant groups. For a given health scenario, the SG was generally
509 scored higher (indicated a more preferred health state) compared the TTO and RS.
510 Some differences between participant groups emerged in the scenario on depression.
- 511 4. Participants ranked the scenarios comparably demonstrating equivalence in
512 discrimination and weighting of the scenarios.
- 513 5. The TTO was preferred over the SG.

514

515 Searching the literature we were unable to locate previous examples where mental health
516 state scenarios were co-produced with service users and carers for use within an e-survey.
517 In line with previous research [8], we found significant differences between the utility scores
518 when using different types of utility measure within each scenario. Similar patterns to those

519 found in other studies were identified, with respondents scoring the highest utility when using
520 the SG and lowest utility when using the RS methodology [27].
521 Of particular interest, service users gave a lower utility value (indicating a less preferred
522 health state) for the depression scenario than healthcare professionals and interested members
523 of the public. Isacson et al. (2005) found that people with depression rated their health state
524 utilities significantly lower than those without [28]. The literature to date suggests that “well-
525 informed” respondents (i.e., people who have experienced the condition) may score the
526 scenario as less threatening and therefore give a higher utility score than respondents who did
527 not share that experience. This is the converse of what is seen in this data, and therefore does
528 not fit with theories such as the disability paradox [29] or the stress-appraisal-coping
529 paradigm [30]. It is important to acknowledge that we do not know which of our respondents
530 had experienced depression and therefore it is unclear whether these findings are due to direct
531 experiences or knowledge relating to a hypothetical health state. Stiggelbout (2008) provides
532 a thorough review on how scenarios are interpreted and judgements are made by people with
533 lived experience and those naïve to the lived experience during the process of scoring utilities
534 [4]. Of particular interest to the field of mental health is the focus of the illness in the
535 person’s life and their constructed meaning. One study showed a recovery-focused approach
536 to interpreting the illness where people with the human immunodeficiency virus (HIV)
537 reframed living with the illness positively by focusing on how HIV fit in with the broader
538 context of their life rather than purely focusing on the impact on their health [31].

539 **Ranking**

540 With regard to how the scenarios were ranked, there was consistency across participant
541 groups in ranking the psychosis scenario as the most undesirable scenario to live with. This
542 was irrespective of utility measure used and it may have implications for service users

543 prioritising treatments that could maximize preferences or health gain. Of course,
544 prioritisation will depend on the estimated gain from any actual intervention.

545 **Acceptability**

546 **Completion of the survey**

547 In this pilot study a high proportion of service users, healthcare professionals and members of
548 the public successfully completed the utility scores for five described healthcare scenarios.
549 There were no substantial differences demographically between participants who completed
550 the survey and those that did not.

551 **Difficulty**

552 In terms of difficulty in completing the online survey, we found that there was a suggestion
553 that healthcare professionals and members of the general public perceive the utility measures
554 as more difficult than service users. However there were no significant between-group
555 differences. Arguably this demonstrates that mental health service users are just as capable of
556 scoring utility scenarios as are members of the public and healthcare professionals. However,
557 this interpretation should be treated cautiously as the qualitative data suggests that there is
558 some difficulties with the SG and TTO utility methods for all participant groups. Participants
559 found the scenarios and scoring mechanisms difficult to understand and were uncertain over
560 how to score the scenarios overall. Some respondents also had concerns around accepting the
561 philosophical notions of trading, 'giving up life years' or 'risking', indicating that face
562 validity within the scenarios remains a challenge. Consistent across the groups was a
563 preference for scoring the TTO. Participants found the TTO easier to understand as they were
564 able to relate to losing years of life more readily than accepting an increased risk of death
565 (SG).The literature also suggests that the TTO is preferred by some for the relative ease of

566 use compared to the SG and has been reported as consistent with individual preferences [9,
567 12] and the most frequently used method [14].

568

569 **Zero-traders**

570 Only 5% of participants using both TTO and the SG for psychosis refused to trade. Zero-
571 traders were most prevalent in the scenario for medication side effects with 20% refusing to
572 trade time (TTO) and 22% unwilling to gamble on an increased risk of death (SG). This may
573 be a function of participants accepting the side effects as a necessary albeit an undesirable
574 aspect of treatment.

575

576 **Limitations**

577 Some participants started the survey but did not complete it (non-completers). The reason for
578 this is unknown, although feedback from other respondents suggest it may have been due to
579 the format of the survey and complexity of questions.

580 Recruitment was voluntary using an internet link, some degree of self-selection bias is likely
581 and probably resulted in a less representative sample. Table 1 indicates that the sample of
582 service users is unusually well-educated, with 37% having a higher degree and another 22% a
583 degree. This is higher than the average in the UK, where 34.4% of the population is estimated
584 to have achieved a degree-level qualification or above [32]. Given the nature of the research
585 it may not be surprising that the sample is relatively well educated, and does limit the
586 generalizability of the findings. The use of online surveys can also pose a challenge for
587 people who do not have access to a computer and this may have had an impact on
588 recruitment. However it is difficult to estimate the true impact of any potential selection bias
589 when data on non-participants is unavailable [33].

590

591 Respondents identified themselves as being healthcare professionals, members of the public,
592 service users or carers, responses which cannot be verified by the researchers. Additionally
593 for many participants these categories are not exclusive and there will be overlap with people
594 identifying with more than one category. For those who identified as service users we have
595 no information about their clinical condition (e.g. diagnosis, severity, duration of time living
596 with the condition) and therefore associations with scoring disease specific-scenarios was not
597 possible [17]. Because we did not use quality of life measures alongside the utility measures
598 convergent validity could not be assessed.

599 In this study we did not control for the order effects of scenario presentation and the potential
600 that scores were moderated by anchoring.

601

602 There are also unresolved questions about how to measure health. Our measures informed by
603 our co-produced scenarios also included an element of social participation. This is an
604 important consideration when proposing to measure mental health with people whose
605 condition, recovery [34] and quality of life is affected by broader social considerations such
606 as housing or employment experiences and interventions. One of our scenarios included the
607 description of neighbours' perceptions of the person with the illness, which may reflect real
608 issues around relating to inclusiveness, stigma and even the reality of experiencing paranoia,
609 but here there is also a danger of stretching the concept of social participation.

610

611 It may be argued that these types of analyses lend themselves more to moderate disorders, the
612 treatment for which is typically designed to ameliorate symptoms as part of improving mental
613 health. With a recovery model in prominence [34], symptom control may not always be the
614 sole concern for severe and enduring disorders such as schizophrenia, where many

615 interventions would seek to target quality of life much more broadly (including housing,
616 employment and other measures of recovery and social participation). Biases in the direction
617 of understating the benefits of these factors on the quality of life of individuals could arise
618 and this may be a further limitation.

619

620 In addition, some conditions are not susceptible to adaptation, and they interrupt daily life
621 almost continually. By their very nature they draw attention to themselves (one cannot just
622 think about something else most of the time); for example, the pre-occupative nature of
623 depression or chronic pain. With this in mind, service users may give depression a lower
624 utility score (i.e. less preferred health state) than other groups, with a risk thereafter for
625 utility weights to be given that are too high. Additionally every preference elicitation
626 question, by their nature, focuses our attention on something, and so we will generally be led
627 to overstate the relative importance to our lives of the things that we are asked to focus on
628 [35]. For equal consideration is the evidence that suggests that the strength of preference
629 may also be a poor guide to the intensity of experience [36-38] and a propensity for us to
630 exaggerate the extent to which we will attend to the state being valued (Dolan and
631 Kahneman, 2008), with us all being members of the ‘public’ and ‘patients’ and therefore
632 susceptible to exaggeration [35].

633 Dolan et al. (2010) also point out, trade-off responses themselves are related to the frequency
634 and intensity of negative thoughts about health in ways that may not have been previously
635 well captured by any of the proposed valuation methods [35]. Different values may also
636 capture “experience” rather than “preference.” Dolan and Kahneman (2008) cite Smith et
637 al’s (2006) work with the following example; a patient with a colostomy thinks they are
638 happy with the colostomy, and expects to be happy again without it. However, when it is
639 removed they remember their previous state (of having the colostomy) as being unacceptable

640 and, in terms of preferences, they report that they would be willing to pay a great deal,
641 including life-years, to get rid of that state [39], a reflection of the extremely negative prior
642 experience.

643

644 Lastly it is worth considering in the treatment of mental health preferences the potential for
645 ‘cognitive denial’, where patients may find it difficult to admit how poor their health really is,
646 or ‘suppressed recognition of full health’ where patients cease to realize what full health may
647 be like and have ‘lowered expectation’ overall. [40].

648 **Future development of utility measures**

649

650 Future research of this type may provide a more rigorous assessment of how health is being
651 conceptualized in the development of such scenarios, while finding ways of helping to create
652 scenarios and scoring mechanisms that are less complex. Possibly a greater challenge will be
653 to create scenarios that do not lead to philosophical objections.

654

655 Nord et al. [41] discuss the use of QALYs in terms of ex ante and ex post. Ex ante is the
656 more traditional approach and refers to health utility judgements made by the general public
657 from behind a ‘veil of ignorance’. There is merit in the ex post approach which refers to the
658 utilization of direct experience of the health state as “experienced utility”. The participants
659 within this study are a combination of both ex ante and ex post participants. It may be
660 beneficial to identify previous health experiences in respondents but conversely it may
661 influence willingness to participate if scenarios are felt less ‘hypothetical.’

662

663 **Conclusion**

664 This study involved service users and reports the initial steps towards developing and
665 embracing a process of research co-production in a complex field [42, 43]. Additional
666 studies involving service users in utility measurement are needed in the attempt to promote
667 sensitive measurement design, increase instrument validity, study feasibility and the
668 acceptability of the measures. Future studies may aim to build on more extensive
669 involvement by developing knowledge and understanding to include service users in the
670 analysis of data and interpretation of results [44].

671

672 Traditionally there have been wide variations in the utility values reported contributing to an
673 overall lack of clarity in reporting methods used to elicit the utility values [45]. This study
674 offers data to compare different valuation methods in order to help assess their feasibility
675 whilst at the same time transparently reporting the methods and some of the difficulties and
676 limitations of our approach. It adds to the limited qualitative evidence reported alongside
677 utility scores for a range of health states and offers insights into factors that influenced
678 respondents' decisions, the relative difficulty of and preferences for measures used. This will
679 help inform our future research and that of others to better prepare such utility design in the
680 future.

681

682 **Acknowledgements**

683 We thank the participants of this study for taking the time to make contributions and Rethink Mental
684 Illness for supporting the promotion of the study. We particularly thank members of the SUGAR
685 group for shaping this research. Membership of SUGAR at the time of this study is listed below, with
686 their lead author and contact person:

687

688 Nelly Adongakulu, Mike Ahern, Mary Amuda, Mandy Bannister, Claudette Brandon, Fadeke Coker,
689 June Hanshaw, Jay Hudson, Richard Humm, Jagadish Jha, Marybel Moore, Isaac Samuels, Jean
690 Taylor.

691 All SUGAR members are Honorary Research Fellows, Centre for Mental Health Research, School of
692 Health Sciences. Lead author for the group and contact email address: Richard Humm
693 richardhumm@hotmail.com

694

695

696

698 **References**

699

- 700 1. Vos T, Barber RM, Bell B, Bertozzi-Villa A, Biryukov S, Bolliger I et al. : **Global, regional,**
701 **and national incidence, prevalence, and years lived with disability for 301 acute and**
702 **chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the**
703 **Global Burden of Disease Study 2013.** *The Lancet* 2015, **386**(9995):743-800.
- 704 2. Davis S: **Annual Report of the Chief Medical Officer 2013: Public Mental Health**
705 **Priorities-Investing in the Evidence.** London: *Department of Health* 2014.
- 706 3. **Economic and social costs of mental health problems in 2009/10**
- 707 4. Stiggelbout AM, Vogel Voogt D: **Health state utilities: a framework for studying the gap**
708 **between the imagined and the real.** *Value in Health* 2008, **11**(1):76-87.
- 709 5. Dolan P, Kahneman D: **Interpretations of utility and their implications for the valuation**
710 **of health.** *The Economic Journal* 2008, **118**(525):215-234.
- 711 6. Gamelin A, Sastre MTM, Sorum PC, Mullet E: **Eliciting utilities using functional**
712 **methodology: People’s disutilities for the adverse outcomes of cardiopulmonary**
713 **resuscitation.** *Quality of life research* 2006, **15**(3):429-439.
- 714 7. Doctor JN, Bleichrodt H, Lin HJ: **Health utility bias: a systematic review and meta-**
715 **analytic evaluation.** *Medical Decision Making* 2008.
- 716 8. Wee HL, Li SC, Xie F, Zhang XH, Luo N, Feeny D et al.: **Validity, Feasibility and**
717 **Acceptability of Time Trade-Off and Standard Gamble Assessments in Health**
718 **Valuation Studies: A Study in a Multiethnic Asian Population in Singapore.** *Value in*
719 *health* 2008, **11**(s1):S3-S10.
- 720 9. Bleichrodt H: **A new explanation for the difference between time trade-off utilities and**
721 **standard gamble utilities.** *Health economics* 2002, **11**(5):447-456.
- 722 10. Elkin EB, Cowen ME, Cahill D, Steffel M, Kattan MW: **Preference assessment method**
723 **affects decision-analytic recommendations: a prostate cancer treatment example.**
724 *Medical Decision Making* 2004, **24**(5):504-510.
- 725 11. Bennett KJ, Torrance GW, Boyle MH, Guscott R: **Cost-utility analysis in depression: the**
726 **McSad utility measure for depression health states.** *Psychiatric Services* 2000.
- 727 12. Morimoto T, Fukui T: **Utilities measured by rating scale, time trade-off, and standard**
728 **gamble: review and reference for health care professionals.** *Journal of Epidemiology*
729 2002, **12**(2):160-178.
- 730 13. Gold MR SJ, Russell LB, Weinstein MC. : **Cost-Effectiveness in Health and Medicine.** New
731 York: Oxford University Press; 1996.
- 732 14. Brazier J: **Measuring and valuing mental health for use in economic evaluation.** *Journal*
733 *of health services research & policy* 2008, **13**(suppl 3):70-75.
- 734 15. König H-H, Günther OH, Angermeyer MC, Roick C: **Utility Assessment in Patients with**
735 **Mental Disorders.** *Pharmacoeconomics* 2009, **27**(5):405-419.
- 736 16. Flood C: **Should " standard gamble" and "time trade off" utility measurement be used**
737 **more in mental health research?** *The journal of mental health policy and economics* 2010,
738 **13**(2):65-72.
- 739 17. Voruganti LN, Awad AG, Oyewumi LK, Cortese L, Zirul S, Dhawan R: **Assessing health**
740 **utilities in schizophrenia.** *Pharmacoeconomics* 2000, **17**(3):273-286.
- 741 18. Briggs A, Wild D, Lees M, Reaney M, Dursun S, Parry D et al.: **Impact of schizophrenia**
742 **and schizophrenia treatment-related adverse events on quality of life: direct utility**
743 **elicitation.** *Health and quality of life outcomes* 2008, **6**(1):1.

- 744 19. Lenert L, Morss S, Goldstein MK, Bergen M, Faustman W, Garber AM: **Measurement of**
745 **the validity of utility elicitation performed by computerized interview.** *Medical care*
746 1997, **35**(9):915-920.
- 747 20. Revicki DA, Hanlon J, Martin S, Gyulai L, Ghaemi SN, Lynch F et al.: **Patient-based**
748 **utilities for bipolar disorder-related health states.** *Journal of affective disorders* 2005,
749 **87**(2):203-210.
- 750 21. Konig H-H: **Measuring preferences of psychiatric patients-A review of the use of**
751 **standard gamble, time trade-off and contingent valuation in patients with depression or**
752 **schizophrenia.** *Psychiatrische Praxis* 2004, **31**(3):118-127.
- 753 22. Papageorgiou K, Vermeulen KM, Leijten FR, Buskens E, Ranchor AV, Schroevers MJ:
754 **Valuation of depression co-occurring with a somatic condition: feasibility of the time**
755 **trade-off task.** *Health Expectations* 2015, **18**(6):3147-3159.
- 756 23. Simpson A, Jones J, Barlow S, Cox L: **Adding SUGAR: service user and carer**
757 **collaboration in mental health nursing research.** *Journal of psychosocial nursing and*
758 *mental health services* 2013, **52**(1):22-30.
- 759 24. Eysenbach G: **Improving the quality of Web surveys: the Checklist for Reporting Results**
760 **of Internet E-Surveys (CHERRIES).** *Journal of medical Internet research* 2004, **6**(3):e34.
- 761 25. Corp I: **SPSS statistics for Windows. Version 21.0.** In.: Author Armonk, NY; 2012.
- 762 26. Braun V, Clarke V: **Using thematic analysis in psychology.** *Qualitative research in*
763 *psychology* 2006, **3**(2):77-101.
- 764 27. Arnold D, Girling A, Stevens A, Lilford R: **Comparison of direct and indirect methods of**
765 **estimating health state utilities for resource allocation: review and empirical analysis.**
766 *Bmj* 2009, **339**.
- 767 28. Isacson D, Binglefors K, von Knorring L: **The impact of depression is unevenly distributed**
768 **in the population.** *European Psychiatry* 2005, **20**(3):205-212.
- 769 29. Ubel PA, Loewenstein G, Schwarz N, Smith D: **Misimagining the unimaginable: the**
770 **disability paradox and health care decision making.** *Health Psychology* 2005, **24**(4S):S57.
- 771 30. Lazarus RS, Folkman S: **Stress, appraisal, and coping:** Springer publishing company; 1984.
- 772 31. Tsevat J, Sherman SN, McElwee JA, Mandell KL, Simbartl LA, Sonnenberg FA et al.: **The**
773 **will to live among HIV-infected patients.** *Annals of internal medicine* 1999, **131**(3):194-
774 198.
- 775 32. Ball C: **Most people in the UK do not go to university - and maybe never will.** In: *The*
776 *Guardian.* [www.theguardian.com/higher-education-network/blog/2013/jun/04/higher-](http://www.theguardian.com/higher-education-network/blog/2013/jun/04/higher-education-participation-data-analysis)
777 [education-participation-data-analysis](http://www.theguardian.com/higher-education-network/blog/2013/jun/04/higher-education-participation-data-analysis); 2013.
- 778 33. Bethlehem J: **Selection bias in web surveys.** *International Statistical Review* 2010,
779 **78**(2):161-188.
- 780 34. Anthony WA: **Recovery from mental illness: The guiding vision of the mental health**
781 **service system in the 1990s.** *Psychosocial rehabilitation journal* 1993, **16**(4):11.
- 782 35. Dolan P: **Thinking about it: thoughts about health and valuing QALYs.** *Health*
783 *Economics* 2011, **20**(12):1407-1416.
- 784 36. Schkade DA, Kahneman D: **Does living in California make people happy? A focusing**
785 **illusion in judgments of life satisfaction.** *Psychological Science* 1998, **9**(5):340-346.
- 786 37. Wilson TD, Gilbert DT: **Affective forecasting.** *Advances in experimental social psychology*
787 2003, **35**(35):345-411.
- 788 38. De Wit GA, Busschbach JJ, De Charro FT: **Sensitivity and perspective in the valuation of**
789 **health status: whose values count?** *Health economics* 2000, **9**(2):109-126.
- 790 39. Smith DM, Sherriff RL, Damschroder L, Loewenstein G, Ubel PA: **Misremembering**
791 **colostomies? Former patients give lower utility ratings than do current patients.** *Health*
792 *Psychology* 2006, **25**(6):688.
- 793 40. Versteegh M, Brouwer W: **Patient and general public preferences for health states: a call**
794 **to reconsider current guidelines.** *Social Science & Medicine* 2016, **165**:66-74.
- 795 41. Nord E, Daniels N, Kamlet M: **QALYs: some challenges.** *Value in Health* 2009, **12**(s1):S10-
796 S15.

- 797 42. Gillard S, Turner K, Lovell K, Norton K, Clarke T, Addicott R et al.: **“Staying native”:**
798 **coproduction in mental health services research.** *International Journal of Public Sector*
799 *Management* 2010, **23**(6):567-577.
- 800 43. Clark M, Pinfold V, Szymczynska P, Hamilton S, Peacocke R, Dean S, et al. **Co-production**
801 **in mental health research: reflections from the People Study.** *Mental Health Review*
802 *Journal* 2015, **20**(4):220-231.
- 803 44. Gillard S, Simons L, Turner K, Lucock M, Edwards C: **Patient and public involvement in**
804 **the coproduction of knowledge reflection on the analysis of qualitative data in a mental**
805 **health study.** *Qualitative Health Research* 2012, **22**(8):1126-1137.
- 806 45. Mohiuddin S, Payne K: **Utility Values for Adults with Unipolar Depression Systematic**
807 **Review and Meta-Analysis.** *Medical Decision Making* 2014:0272989X14524990.

808

809 **Supporting information**

810 **S15** Table available as information file for the five scenarios used in the utility measurements
811 and offered to participants in the online questionnaire