The University Of Sheffield.

School of
Health
And
Related
Research

Health Economics & Decision Science (HEDS)

Discussion Paper Series

Valuing Health-Related Quality of Life: An EQ-5D-5L Value Set for England

Authors: Nancy Devlin, Koonal Shah, Yan Feng, Brendan Mulhern, Ben van Hout

Corresponding author: Nancy Devlin

Office of Health Economics, London. Southside, 7th Floor, 105 Victoria Street, London, SW1E 6QT
Tel: +44 (0) 20 7040 8858
Email: ndevlin@ohe.org

No. 16.02

Valuing Health-Related Quality of Life: An EQ-5D-5L Value Set for England

Nancy Devlin[1], Koonal Shah[1], Yan Feng[1], Brendan Mulhern[2], Ben van Hout[3]

[1] Office of Health Economics, London

[2] Centre for Health Economics Research and Evaluation, University of Technology Sydney

[3] School of Health and Related Research, University of Sheffield

January 2016

Corresponding author :

Nancy Devlin

ndevlin@ohe.org

Office of Health Economics

Southside, 105 Victoria Street

London SW1E 6QT

United Kingdom

Tel: +44 207 747 8858

Acknowledgements

This study was funded by a Department of Health Policy Research Programme grant (NIHR PRP 070/0073). Additional funding and technical support was provided by the EuroQol Research Foundation. The authors are grateful to the Project Steering Group, chaired by Dr Alan Glanz (Department of Health), and to colleagues from the EuroQol Research Foundation, University of Sheffield and Office of Health Economics, for advice and feedback received throughout this study. We are particularly grateful to John Brazier for his comments on an earlier draft.

Disclaimers

Views expressed in the paper are those of the authors, and are not necessarily those of the Department of Health or the EuroQol Research Foundation.

This Discussion Paper reports our most up-to-date analyses in order to make this research publicly accessible and to stimulate discussion and critical comment. The value set reported in this Discussion Paper has not been approved or endorsed by any external bodies. It should be noted that this version, and any other versions made available ahead of journal publication, necessarily have interim status as the peer review process may necessitate changes to the analyses and results. Any use of the content of this paper is the sole responsibility of the user. The authors assume no responsibility for, and expressly disclaim all liability for, any consequences resulting from the use of the information herein.

This HEDS Discussion Paper also has been published as Office of Health Economics Research Paper 16/01.

Contents

ABSTRACT

Objectives: Measures of patient-reported health are increasingly used in clinical and health system decisions, and the EQ-5D is one of the most widely used questionnaires. It is recommended by NICE and is widely used in clinical trials, as well as in population health surveys and the NHS PROMs programme. A new version, the EQ-5D-5L, is now available. The objective of this study is to establish how important different sorts of health problems are to overall quality of life, and to produce the set of scores ('value set') required to use EQ-5D-5L data in decision-making and priority setting in the English NHS.

Design: The study design followed an international research protocol. Each participant valued 10 health states using a time trade-off approach and completed seven discrete choice tasks. The data are used to model values for all 3,125 states described by the EQ-5D-5L.

Setting: England general population.

Participants: Data were collected in face-to-face interviews with 996 adult members of the general public, selected at random from residential postcodes. The sample is broadly representative of the general population.

Results: The data obtained from participants had good face validity. Problems with pain/discomfort and anxiety/depression were the most important factor in overall quality of life. Values ranged from -0.281 (for extreme problems on all dimensions) to 0.951.

Conclusions: The value set reported here will have important implications for public decisions made using EQ-5D-5L data. There are considerably fewer states judged to be 'worse than dead' compared to the current EQ-5D value set (4.93%, compared with over one-third) and the minimum value is higher (-0.281 compared to -0.594). The results imply that QALY gains for interventions seeking to improve very poor health will be smaller using the EQ-5D-5L tariff, and may previously have been overestimated.

1. INTRODUCTION

Clinical practice requires decisions to be made under uncertainty, whereby any decision may have a range of different outcomes. To make the 'best' decision, potential outcomes need ordering and valuing. Such decisions are made both at the individual level, such as choosing the optimal treatment for a patient, and at the national level, such as choosing how to allocate resources between treatments for different patient groups and across different health conditions.

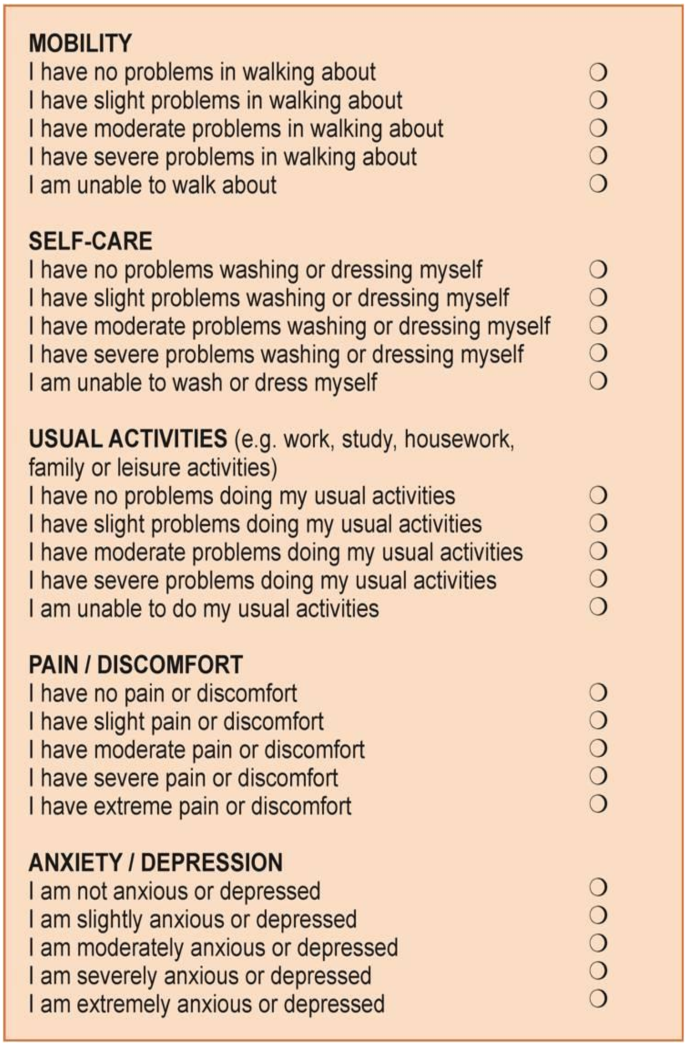
Clinical decisions often affect patients' health-related quality of life (HRQL). Evidence on patients' HQRL can be obtained using 'patient-reported outcome' (PRO) measures. These may be condition-specific or generic. Condition-specific PROs focus on specific health problems and aim to provide detailed information about the impacts of the condition, disregarding problems which are atypical for the condition. Generic PROs aim to cover a more general spectrum of health problems, and are designed to be applicable for any health condition. They can capture co-morbidities, and allow comparisons with 'population norms'. Evidence obtained from generic measures can be used to compare both the impact of health problems and the benefits offered by treatments across different patient populations and disease areas. This makes these data particularly useful for the decisions made by general practitioners, commissioners, regulators, the National Institute for Health and Care Excellence (NICE) and the NHS.

The EQ-5D is the most widely used generic PRO questionnaire internationally (Kind et al., 2005). It is the instrument recommended by NICE for evidence submitted to its technology appraisal process (NICE, 2013). It has also proved useful in population health surveys and in the English NHS PROMs programme (Devlin and Appleby 2010). The EQ-5D asks patients to indicate whether they have no, some or extreme problems on each of five dimensions of health: mobility; self-care; usual activities; pain/discomfort; anxiety/depression.

The EQ-5D is a valid and reliable measure in many disease areas (Janssen et al., 2011; Wailoo et al., 2010; Pickard et al., 2007). However, there have been concerns that three response options for each dimension may not adequately capture milder health problems experienced by patients, and smaller changes between different health states. A new version of the instrument, the EQ-5D-5L, was developed to improve sensitivity and to standardise the language used across dimensions (Herdman et al., 2011). The EQ-5D-5L comprises the same five dimensions, but increases the available response options ('levels') from three to five (no; slight; moderate; severe; extreme problems/unable to)

– see Figure 1. The five dimensions and five levels of the EQ-5D-5L describe 3,125 (5^5) unique health states, compared to the 243 (3^5) described by the EQ-5D.

Figure 1. The EQ-5D-5L descriptive system



MOBILITY
I have no problems in walking about ○
I have slight problems in walking about ○
I have moderate problems in walking about ○
I have severe problems in walking about ○
I am unable to walk about ○

SELF-CARE
I have no problems washing or dressing myself ○
I have slight problems washing or dressing myself ○
I have moderate problems washing or dressing myself ○
I have severe problems washing or dressing myself ○
I am unable to wash or dress myself ○

USUAL ACTIVITIES (e.g. work, study, housework, family or leisure activities)
I have no problems doing my usual activities ○
I have slight problems doing my usual activities ○
I have moderate problems doing my usual activities ○
I have severe problems doing my usual activities ○
I am unable to do my usual activities ○

PAIN / DISCOMFORT
I have no pain or discomfort ○
I have slight pain or discomfort ○
I have moderate pain or discomfort ○
I have severe pain or discomfort ○
I have extreme pain or discomfort ○

ANXIETY / DEPRESSION
I am not anxious or depressed ○
I am slightly anxious or depressed ○
I am moderately anxious or depressed ○
I am severely anxious or depressed ○
I am extremely anxious or depressed ○

Note: The figure above is provided for illustrative purposes only. The full EQ-5D-5L questionnaire is available from the EuroQol Research Foundation.

The EQ-5D-5L is rapidly being incorporated into routine data collection in clinical settings, clinical trials and population health surveys (such as the GP Patient Survey and the Health Survey for England). It is also used in local initiatives – for example, the Cambridgeshire Community Services NHS Trust collects EQ-5D-5L data to evaluate outcomes from rehabilitation services (Cambridgeshire Community Services NHS Trust, 2013-2014). Its design accounts for the need for a direct link between the measurement and valuation of health, whereby every 'health state' – i.e. combination of health problems – which patients might report on the EQ-5D-5L instrument can be summarised by a single value. In order to be used in the calculation of quality-adjusted life years (QALYs; a metric used in cost-utility analysis that combines survival and HRQL), these health state values needs to summarise how good or bad the health problems described are on a scale anchored at 1 (full health) and 0 (a state equivalent to dead). Health states considered to be worse than dead are given values less than 0. The values are based on the views of the general public who are asked to imagine living in various health states, and to respond to a series of structured questions designed to find out the importance to them of different aspects of health. This approach follows the requirements of NICE (2013) and similar organisations for the use of EQ-5D-5L data in decision making, and reflects a belief that it is the views of the general public – as taxpayers and potential users of health care – that should count, rather than simply those of patients (Gold et al., 1996).

Value sets for the (three-level) EQ-5D are available for a range of countries (Szende et al., 2007). The current UK EQ-5D value set (Dolan, 1997) has values which range from 1 for no problems on any dimension to -0.594 for the worst health state (level 3 problems on each dimension). A number of limitations have been noted with that value set. Among these are that approximately a third of health states described by the EQ-5D were assigned negative values, meaning those health states are valued as being 'worse than dead'. The UK values are rather unique in this respect: all other countries have higher values. Additionally – although also common to other countries – any change in health away from full health to 'some' problems, on any aspect of health, results in a large fall in the overall value (of at least 0.12) on the 0 to 1 scale. These issues with the current EQ-5D value set may have important implications for decisions being made in the NHS. For example, NICE estimates of QALY gains from new treatments may be biased upwards. (Devlin et al., 2011; Devlin et al., 2012).

To date there have been no values specific to the (five-level) EQ-5D-5L available to summarise patients' data. Research has established the relationship between patients' self-reported health on the EQ-5D and on the EQ-5D-5L, enabling EQ-5D-5L data collected from patients to be summarised using the EQ-5D value set, via a 'mapping algorithm' (van Hout et al., 2012) This provides an interim means of scoring EQ-5D-5L data, but perpetuates the limitations of the EQ-5D value set (Devlin et al., 2011; Devlin et al., 2012; Tilling et al., 2010)

The aim of this study is to produce a 'value set' for the EQ-5D-5L that can be used to support decision making in the English NHS. This is one of the first value sets to be reported for the EQ-5D-5L internationally. The study is relevant to clinicians collecting PRO data from patients, and to those using PRO data in health care decisions. It demonstrates the relative importance placed on different types of health problems by people in England – and how that should be reflected in priority-setting.

2. METHODS

The research design and data collection followed a research protocol developed by the EuroQol Research Foundation, a not-for-profit international network of multidisciplinary researchers. The protocol was informed by an extensive programme of methodological research investigating methods for valuing EQ-5D-5L health states (Oppe et al., 2014). Our study was one of the first to use the protocol, and comparable studies are now underway worldwide.

2.1 Methods of eliciting preferences

The study used the EuroQol Valuation Technology (EQ-VT) software, developed specifically for EQ-5D-5L value set studies and administered using computer-assisted personal interviews. Two stated preference methods were used to elicit preferences: time trade-off (TTO) – an approach used in previous EQ-5D valuation studies (Oppe et al., 2014) and accepted by NICE as a 'choice-based' approach (NICE, 2013) and discrete choice experiments (DCE) – an approach that is increasingly used to assess preferences for health states because of the relative simplicity of the tasks (Ryan et al., 2008). The two methods generate different and complementary preference data. TTO elicits a value for each state with 1 and 0 defined as anchor points, whereas DCE generates binary data which allow for the derivation of a scale of non-anchored relative values.

Each interview consisted of the following tasks (in order): self-reported health using EQ-5D-5L, self-reported health on a 0-100 visual analogue scale (EQ-VAS); basic background questions; a practice TTO task (involving the valuation of a simple health state describing being confined to a wheelchair); 10 TTO tasks; structured feedback questions regarding the TTO tasks; seven DCE tasks; structured DCE feedback questions; an (optional) open-ended comment box; and further background questions.

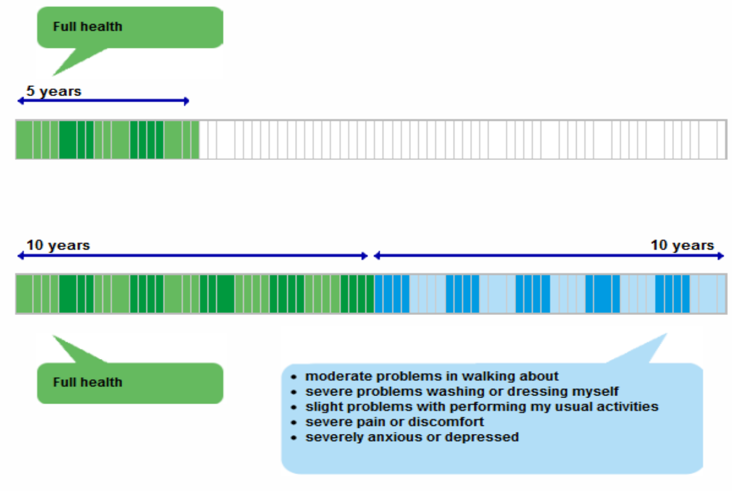
In the TTO tasks, a composite approach was used which involved starting with the 'conventional' TTO (Brazier et al., 2007) for all health states, and shifting to a 'lead time' TTO when participants indicated that they considered the health state to be worse than dead (Devlin et al., 2011; Devlin et al., 2012; Robinson and Spencer, 2006). The composite TTO approach is illustrated in Figures 2a and 2b. Evidence supporting this approach is reported by Janssen et al. (2013).

Figure 2a illustrates the TTO task for health states better than dead (i.e. those with a value between 0 and 1). The participant is asked to imagine living for 10 years from today in a given EQ-5D-5L state, followed by death ('Life B'). The participant's value for that health state is then derived by identifying, using a systematic iterative process, the number of years in full health between 0 and 10 ('Life A') they consider equivalent to that.

The more severe the health state described in Life B, the more years of full health the participant is assumed to be willing to give up in Life A to avoid Life B. For very poor health states, all of the time in Life A may be traded off, indicating that the value for the state is less than or equal to 0. Where this occurs, additional time in full health ('lead time') is added to both Life A and Life B – see Figure 2b. This allows participants to trade off more time, reflecting how much worse than dead they consider the health state to be (within the boundaries of the scale imposed by the task) (Devlin et al., 2012).

Figure 2a. TTO valuation of health states better than dead (i.e. values ≥ 0)

Figure 2b. TTO valuation of health states worse than dead (i.e. values ≤ 0)



Source: Reproduced with the permission of the EuroQol Research Foundation.

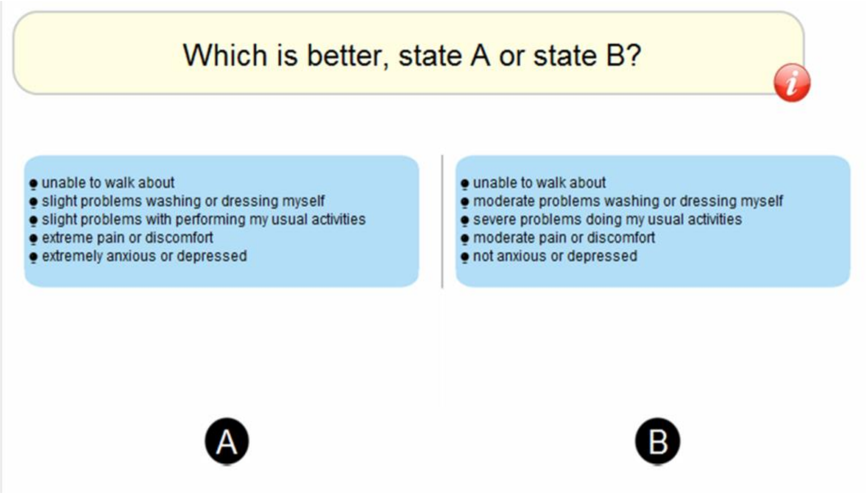
The variant of lead time TTO used in this study involved a 20 year time frame (10 years of lead time followed by 10 years in the health state under evaluation), yielding a minimum value of -1. No additional trade-off questions were asked of those who 'exhausted' their lead time (Devlin et al., 2012).

The iterative procedure used to seek the point of indifference was based on an adaptation of that used in the UK EQ-5D value set study (Dolan 1997). Further details about EQ-VT and the iterative process used in the TTO tasks are provided by Oppe et al (2014).

Each TTO task ends when the participant indicates that they consider Life A and Life B to be 'about the same'. At this point of indifference, the implied value for health states better than dead is calculated by dividing the total number of years in Life A (t) by 10 (the total number of years in Life B). This can be expressed as $V=t/10$, where V is health state value. For example, at the point of indifference shown in Figure 2a, the health state value would be 5/10=0.5. The implied value for health states worse than dead is calculated by subtracting 10 (the number of years of lead time) from the total number of years in Life A, then dividing by 10 (the total number of years in Life B, minus the number of years of lead time). This can be expressed as $V=(t-10)/10$. The point of indifference shown in Figure 2b would suggest a value of (5-10)/10=-0.5. The maximum value is 1, achieved when the participant considers 10 years in the health state to be as good as 10 years in full health. A value of 0 is given when the respondent considers the health state to be no better and no worse than dead. The minimum score of -1 (where all of the lead time is exhausted) is given when the participant considers the prospect of living for 10 years in full health followed by 10 years in the health state to be worse than or equivalent to a life lasting 0 years (i.e. dying now).

In each DCE task (Figure 3), participants were presented with a pair of health states (labelled A and B), with no reference to the duration of the states, and asked to indicate which they considered to be ‘better’ by clicking the appropriate button. No indifference option was included.

Figure 3. Discrete choice experiment task



Which is better, state A or state B?

- unable to walk about
- slight problems washing or dressing myself
- slight problems with performing my usual activities
- extreme pain or discomfort
- extremely anxious or depressed

- unable to walk about
- moderate problems washing or dressing myself
- severe problems doing my usual activities
- moderate pain or discomfort
- not anxious or depressed

Source: Reproduced with the permission of the EuroQol Research Foundation.

2.2 Study design

Overall, 86 health states were selected for valuation via TTO, and 196 pairs of health states for valuation via DCE. Participants were randomly assigned to one of 10 blocks of 10 TTO tasks and to one of 28 blocks of seven DCE tasks. The order in which the states appeared within each type of task was also randomised. None of the pairs in the DCE design included a health state that logically dominated the other (i.e. at least as good on all five dimensions). Each block of TTO tasks contained a combination of mild, moderate and severe health states. All blocks included the worst health state in the EQ-5D-5L descriptive system (described as 55555, i.e. level 5 for all dimensions), and one of the least severe health states (with no problems on four dimensions and slight problems on one dimension, e.g. 21111).

2.3 Data collection

Sample recruitment and interviewing was carried out by the market research company Ipsos MORI. A sample of 2,020 addresses from 66 primary sampling units across

England was randomly selected based on postcode sectors using the Post Office small user Postcode Address File. This includes all private residential accommodation in England (communal establishments, such as prisons and care homes, were excluded). Thirty-seven addresses were selected systematically from an ordered list of all addresses within each sampling unit, ensuring that addresses were spread evenly across it. Interviewers sent an advance letter and information sheet to each selected address, together with an unconditional incentive of six first-class stamps. In each selected dwelling unit, all individuals aged 18 years and over were listed in alphabetical order by first name and one was selected randomly using a selection grid with no substitutes permitted.

If the selected individual gave their informed consent to take part, they were interviewed in their own home. The participant was in control of the computer (laptop) throughout the tasks, with the interviewer guiding them through each step, following a script. The one-to-one setting allowed interviewers to provide detailed instruction and feedback where appropriate (Shah et al., 2013).

Forty-eight interviewers were used, all of whom attended a full day briefing in which they were given intensive training on the methodology and study procedures by the research team. Interim data were monitored at the interviewer level, at least weekly. If an interviewer was found to be generating unusual or poor quality data (defined using criteria based on expected data characteristics, given previous research), they were given additional training. No data were removed at this stage. The study was given approval by the research ethics committee of the University of Sheffield's School of Health and Related Research.

2.4 Piloting

The main study was preceded by a small pilot study (n=49), undertaken in August 2012. The pilot sample was recruited using quotas on age, gender and working status rather than using the systematic approach described above. The aims of the pilot were: to test for technical issues with EQ-VT; to test Ipsos MORI's procedures and methods of encouraging participation; to seek interviewers' feedback on the preliminary script and other materials; to seek participants' feedback on the interview; to examine basic properties of the data generated; and to identify ways to improve the interviewer training process.

The pilot was completed successfully, with no issues with EQ-VT reported by participants or interviewers. Several improvements were made as a result of the feedback received during piloting. These included refinement of the interviewer script and improvements to the interviewer training process.

2.5 Methods of analysis

For both TTO and DCE, a range of descriptive analyses were conducted to assess data characteristics. For TTO this included examining the overall distribution of values, and correlating average values for each health state with its level sum score (a proxy for severity; e.g. the worst health state 55555 has a score of 5+5+5+5+5=25). For DCE, the proportion of participants choosing health state A was assessed in comparison to the difference in level sum scores between A and B, with the implicit assumption that health states with lower level sum scores would be more likely to be chosen overall.

To generate the modelling dataset, we tested a range of possible exclusion criteria, reflecting alternative judgements that might be made about the quality of the data. We wished to minimise exclusions but sought to omit data which were clearly implausible. The final rules for the TTO data were to exclude: (a) participants who gave all 10 health states the same value (all health states cannot plausibly be given the same value given the severity range in each block); and (b) participants who gave the worst state, 55555, a value that was no lower than the value they gave to the mildest health state in their block. Both suggest either misunderstanding or lack of engagement. No DCE data were excluded.

Modelling explicitly addressed observed characteristics of the data (Feng et al., 2015). First, the minimum TTO value is bounded at -1 by design, so we allowed for the possibility of values lower than that using survival analysis approaches for treating censored data. Second, the maximum TTO value is 1, again by design, so although there is an error distribution around observed values, that distribution is necessarily asymmetric at 1, biasing the values for mild health states downwards. Thus, the values at 1 were also considered to be censored. Third, there were some participants who used 0 as the minimum value for more than one health state (almost always including health state 55555). This suggests that these participants were averse to giving negative TTO values. Those values were considered to be censored at zero. Furthermore, some participants gave health state 55555 a value of 0 whilst giving multiple other health states a negative value. This in an example of a ‘logical inconsistency’, since all health states dominate 55555. Those values were censored at 0.

Further, we observed that participants more often disagree about the value of health states that are further away from full health, i.e. the variance of TTO values increases for worse health states. From eyeballing the valuation data at the individual participant level, it was apparent that different groups of participants differed in their use of the scale, resulting in substantially different slopes (i.e. the relationships between disutility and health state severity). This could fundamentally be driven by the heterogeneity of participants in their views about death. The effect of heterogeneity was explored using models that introduced a parameter for the scale of disutility in health, which may differ between participants. The scale of disutility in health was assumed to follow a multinomial distribution with probability density on a number of latent groups. Each of the latent groups has its mean and variance for the distribution of the scale. In our analysis, we assessed models with two, three and four latent groups, eventually judging that the three-group model best fitted the data. We also accounted for heteroskedasticity within each latent group in the model. This was achieved by estimating two parameters per group for modelling the relationship between the mean and the variance of health states.

Models were estimated with different degrees of freedom. The most restrictive model gives different weights to the five dimensions and assumes equal distances between levels. The second most restrictive adds different values for the levels with a distinction between ‘extreme’ and ‘unable to’ (the former is the level 5 label used for the pain/discomfort and anxiety/depression dimensions; the latter is the level 5 label used for the mobility, self-care and usual activities dimensions). The least restrictive model has a parameter for each decrement or step away from ‘no problems’ (level 1) on each dimension. The least restrictive model was considered the best specification as it applied no assumptions about the parameters. In total, the model estimated 20 parameters (4 levels x 5 dimensions). Within these specifications, a range of alternative models was tested to capture the possibility of interaction effects between dimensions and levels.

20 parameter models that involved TTO data showed logical inconsistencies in some dimensions. In the model used to produce the EQ-5D-5L value set for England, we applied restrictions to the parameters. Specifically, the level 2 parameters were estimated first, and parameters for subsequent levels were estimated by adding quadratic terms (which can be non-negative only, thereby ensuring that moving to worse levels of problems always resulted in an increase in disutility).

Models were estimated separately for the TTO and DCE data, and then using a hybrid modelling approach which optimised the likelihood of both data sets at the same time and resulted in a single combined-data value function (Ramos-Goñi et al., in press)

The sociodemographic composition of the sample was checked for representativeness against the general population. Models were estimated on data specific to selected sociodemographic groups (e.g. male participants vs. female participants) to check for any systematic differences.

All analyses were conducted in R3.2.0 and Winbugs 14.

The methods and analyses reported in this paper comply with the CREATE guidelines for reporting valuation studies of multi-attribute utility-based instruments (Xie et al., 2015). The modelling methods are described in greater detail in an accompanying paper by Feng et al. (2015).

3. RESULTS

3.1 Sample

The interviews were conducted between November 2012 and May 2013. Of the individuals invited to take part in the study, 996 completed the valuation questionnaire, comprising TTO tasks, DCE tasks and basic background questions, in full (response rate ≈ 40%). In accordance with the ethical approval for this study, participants who did not complete the valuation questionnaire in full were excluded from the analysis, hence there are no missing TTO and DCE responses in our data set. Full background data were collected for 985 of the 996 participants (98.9%). Table 1 shows that the sample includes a larger proportion of retired individuals and a smaller proportion of younger individuals than in the general population (Office for National Statistics, 2011). The sample also includes a relatively large proportion of individuals with health problems.

Table 1. Background characteristics of the sample

	All participants (n=996)[ii]	**After exclusions (n=912)**[ii]	**General population**[i]
	N (%)	**N (%)**	**%**
Age			
18-29	113 (11.3)	105 (11.5)	20.7%
30-44	298 (29.9)	270 (29.6)	26.3%
45-59	250 (25.1)	227 (24.9)	24.7%
60-74	207 (20.8)	191 (20.9)	18.5%
75+	128 (12.9)	119 (13.0)	9.9%
Gender			
Male	405 (40.7)	372 (40.8)	49.2%
Female	591 (59.3)	540 (59.2)	50.8%
Economic activity			
Employed or self-employed	504 (51.2)	463 (50.8)	59.4%
Retired	278 (28.2)	256 (28.1)	13.1%
Student	20 (2.0)	19 (2.1)	8.8%
Looking after home or family	83 (8.4)	73 (8.0)	4.2%
Long-term sick or disabled	48 (4.9)	42 (4.6)	3.9%
Other / none of the above	52 (5.3)	47 (5.2)	10.6%
Marital status			
Never Married	238 (24.2)	225 (23.4)	34.6%
Married	466 (47.3)	434 (47.6)	46.6%
Same-sex civil partnership	2 (0.20	2 (0.2)	0.2%
Separated [iii]	37 (3.8)	32 (3.5)	2.7%
Divorced	131 (13.3)	119 (13.0)	9.0%
Widowed [iv]	107 (10.9)	99 (10.9)	6.9%

	All participants (n=996)[ii]	After exclusions (n=912)[ii]	General population[i]
Prefer not to say	4 (0.4)	1 (0.1)	N/A
Religion			
Christian	636 (64.6)	575 (63.9)	59.4%
Any other religion	60 (6.0)	53 (5.8)	8.7%
No religion	281 (28.5)	266 (29.6)	24.7%
Religion not stated	8 (0.8)	6 (0.7)	7.2%
Ethnicity			
White	900 (91.4)	832 (92.4)	85.4%
Any other ethnic group	82 (8.3)	67 (7.4)	14.6%
Prefer not to say	3 (0.3)	1 (0.1)	N/A
Day-to-day limitations due to health problem or disability			
Limited a lot	111 (11.3)	95 (10.6)	5.6%[v]
Limited a little	158 (16.0)	144 (16.00	7.1%[v]
Not limited	716 (72.7)	661 973.40	87.3%[v]
Education			
Degree	211 (21.4)	201 (22.3)	
No degree	774 (78.6)	699 (77.7)	
Main language spoken			
English	920 (93.40	847 (94.1)	
Any other language	65 (6.6)	53 (5.9)	
Responsibility for children			
Yes	350 (35.5)	314 (34.90	
No	635 (64.5)	586 (65.1)	
Experience of serious illness			
In self	330 (33.1)	297 (33.0)	
In family	692 (69.5)	636 (70.7)	
In caring for others	416 (41.8)	385 (42.8)	
Self-rated health using EQ-5D-5L			
11111	474 (47.6)	437 948.6)	
Any other health state	522 (52.4)	475 (52.8)	
Self-rated health using EQ-VAS			
<80	334 (33.5)	298 (33.1)	
80-89	256 (25.7)	241 (26.8)	
90-99	337 (33.8)	306 (34.0)	
100	69 (6.9)	67 (7.4)	

[i] Based on results of the 2011 UK Census (Office for National Statistics, 2011), where available

[ii] Data on economic activity, marital status, religion, ethnicity, day-to-day limitations, main language and responsibility for children unavailable for a minority of participants

[iii] Comprises individuals who are separated but still legally married or in a same-sex civil partnership

[iv] Includes individuals who are the surviving partner from a same-sex civil partnership

[v] Census data reported here refers to individuals aged 16-64 only

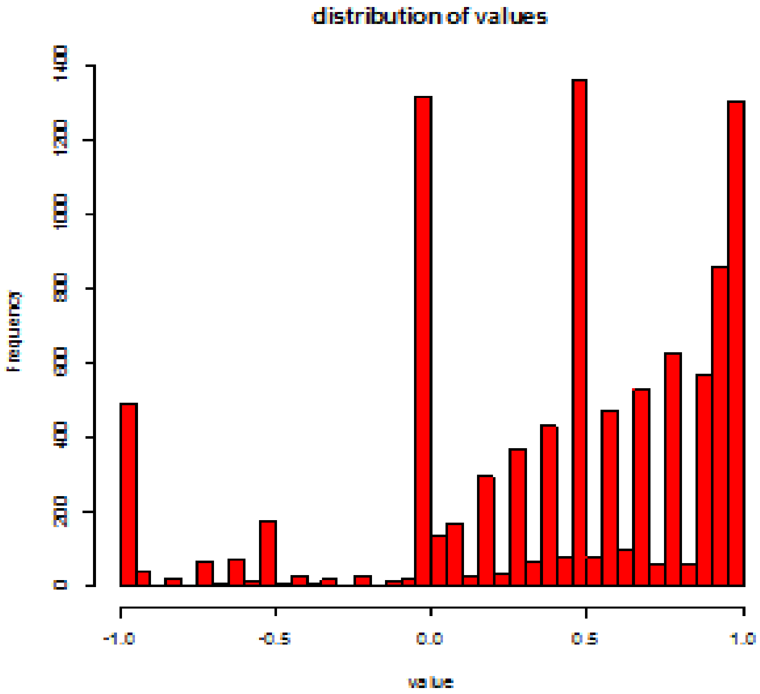
3.2 Descriptive analysis

Figure 4 shows the distribution of observed TTO values. There is some evidence of clustering at key values on the scale (1, 0.5 and 0) and of digit preference (where most values provided are ‘round’ numbers). Health states were given a value of -1 (indicating that the respondents exhausted all of the lead time available to them) on 400 occasions (4.02% of all TTO observations). There are few observations between 0 and -0.5.

We observed evidence of interviewer effects, with different proportions of worse than dead values depending on which interviewer participants were interviewed by.

As a simple test of the face validity of the data, the means and medians of the TTO values were plotted against the level sum scores of the health states. The results (not shown here) show the expected negative relationship, i.e. the worse the health state, the lower the average observed value. Similarly, the proportion of those choosing A or B in the DCE tasks was strongly correlated to the difference in level sum score between the health states, i.e. the greater the difference in severity between any two states, the more likely participants were to choose the health state with the lower level sum score.

Figure 4. Distribution of observed TTO values



3.3 Exclusion criteria

Twenty-three participants (2.3%) gave all 10 health states the same value, and 61 participants (6.1%) valued 55555 no lower than the value they gave to the mildest health state. Excluding these participants gave a core modelling dataset of 912 participants (9,120 TTO observations). Post-exclusions, health states were given a value of -1 on 392 occasions (4.30% of all TTO observations).

Of the remaining individuals, 150 participants with more than one health state valued at 0 were treated as censored on the assumption that 0 was the lowest value they were willing to use. Censoring was also applied to 27 participants with inconsistent negative data (where 55555 was given a value of 0 and more than one other health states were given negative values).

3.4 Modelling results and the value set

When addressing the weights given to the different dimensions, the TTO data suggest weights for mobility, self-care, usual activities, pain/discomfort and anxiety/depression of 0.052, 0.046, 0.053, 0.078 and 0.077, respectively. The DCE data suggest weights of 0.338, 0.241, 0.205, 0.406 and 0.393, respectively. Both methods suggest that pain/discomfort and anxiety/depression should receive the greatest weight.

Table 2 presents the EQ-5D-5L value set based on a hybrid model combining the TTO and DCE data. The minimum value is -0.281 (for the worst health state, 55555) with 4.93% of the 3,125 health states described by the EQ-5D-5L being valued as worse than dead. The size of the coefficients in Table 2 reflects the relative weight placed on different sorts of health problems by our sample. For example, at the worst level of problem (level 5) that can be experienced, pain/discomfort is considered to have the greatest overall impact on HRQL (0.341), followed by anxiety depression (0.301), mobility (0.275), self-care (0.217) and usual activities (0.190). At lower levels of problems, anxiety/depression has the largest effect on HRQL, followed by pain/discomfort, self-care, usual activities and mobility.

Table 2. An EQ-5D-5L value set for England

	Central estimate	Standard Deviation	Value for health state 23245
Constant	1.000		1.000
Mobility			
slight	0.051	0.004	0.051
moderate	0.063	0.004	
severe	0.212	0.006	
unable	0.275	0.006	
Self-care			
slight	0.057	0.004	
moderate	0.076	0.004	0.076
severe	0.181	0.005	
unable	0.217	0.005	
Usual activities			
slight	0.051	0.004	0.051
moderate	0.067	0.004	
severe	0.174	0.005	
unable	0.190	0.005	
Pain/discomfort			
slight	0.060	0.004	
moderate	0.075	0.005	
severe	0.276	0.007	0.276
extreme	0.341	0.008	
Anxiety/depression			
slight	0.079	0.004	
moderate	0.104	0.005	
severe	0.296	0.007	
extreme	0.301	0.007	0.301
Probability (group 1)	0.397	0.019	0.397 x 0.427 + 0.270 x 0.939 + 0.333 x 1.635 =0.9675
Probability (group 2)	0.270	0.018	
Probability (group 3)	0.333	0.018	
Slope (group 1)	0.427	0.031	
Slope (group 2)	0.939	0.067	
Slope (group 3)	1.635	0.017	
The value for health state 23245			1-0.9675 x (0.051+ 0.076+0.051+0.276+0.301) =0.270

Table 2 provides a worked example of how to calculate the values for health state 23245, where the relevant decrement for each level of problem on each dimension is subtracted from the constant. The coefficients from the three latent classes are reported by the three probabilities and three slopes. These six coefficients are used to calculate the weighted average of slope. The weighted average of slope is a fixed coefficient (i.e. 0.9675) to be multiplied by the sum of the five decrements while calculating values for all health states.

3.5 EQ-5D and EQ-5D-5L value set comparisons

Table 3 compares the EQ-5D-5L value set with the original EQ-5D value set (Dolan, 1997) and the crosswalk value algorithm reported by van Hout et al. (2012). The EQ-5D-5L value set has a higher value for the worst possible health state and substantially fewer worse than dead values. The decrement from the best (11111) to next best health state is smaller for the EQ-5D-5L value set, as might be expected given differences in number of levels and labelling between the instruments (e.g. 11211 describes 'slight' problems performing usual activities in the five-level instrument and 'some' problems in the three-level version). Pain/discomfort has the largest decrement, while self-care and usual activities have the smallest. Figure 5 shows that the EQ-5D-5L value set has a normal distribution, in contrast to the EQ-5D value set which was characterised by two peaks. (Parkin et al., 2014)

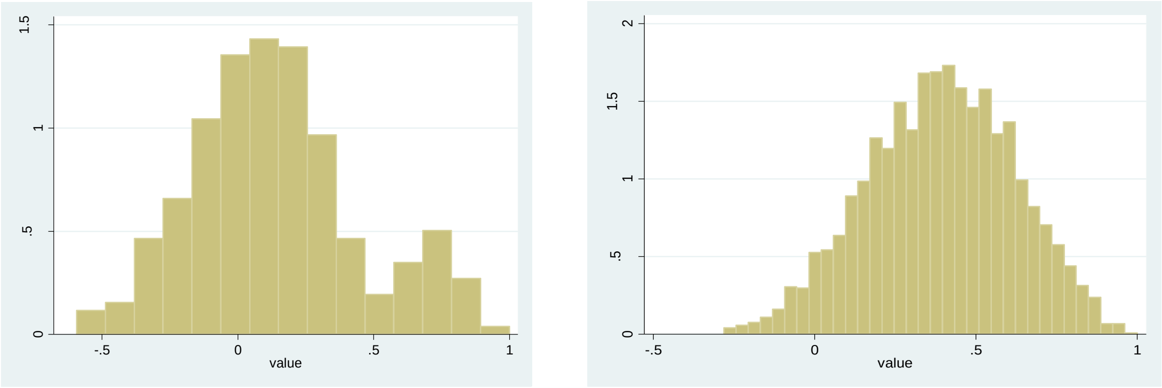
Table 3: Comparison of the key characteristics of 5L values, crosswalk values and 3L values

	EQ-5D-5L value set	**Crosswalk value set**	**EQ-5D value set**
% health states worse than dead	4.93% (154 out of 3,125)	26.7% (833 out of 3,125)	34.6% (84 out of 243)
Preferences regarding dimensions (ordered from most to least important §)	Pain/Discomfort	Pain/Discomfort	Pain/Discomfort
	Anxiety/Depression	Mobility	Mobility
	Mobility	Anxiety/Depression	Anxiety/Depression
	Self-care	Self-care	Self-care
	Usual Activities	Usual Activities	Usual Activities
Value of 55555 (33333)	-0.281	-0.594	-0.594
Value of 11112*	0.924	0.879	0.848
Value of 11121*	0.942	0.837	0.796
Value of 11211*	0.951	0.906	0.883
Value of 12111*	0.945	0.846	0.815
Value of 21111*	0.951	0.877	0.850
Minimum value	-0.281	-0.594	-0.594
Maximum value	1	1	1

*Note that for each of the asterisked health states, the level of problems indicated on the five-level and three-level versions of EQ-5D differ: for example, on the EQ-5D-5L, 11112 means no problems on any dimension except *mild* problems with anxiety/depression, whereas on the EQ-5D, 11112 means no problems on any dimensions except *some* problems with anxiety/depression. *A priori*, we would expect the values for these health states to be higher in the EQ-5D-5L value set than the EQ-5D value set, which is what we observe.

§Importance is judged by the size of the coefficient for level 5 in each dimension.

Figure 5. Frequency of values in the EQ-5D (left) and EQ-5D-5L (right) value sets



4. DISCUSSION

We have reported a value set for the EQ-5D-5L, based on the preferences of a random sample of the English general public. Such value sets promote consistency and comparability in assessments of HRQL across different patient groups. The value set can be used to summarise EQ-5D-5L data collected from patients in a wide range of contexts and in the economic evaluation of health care interventions to support resource allocation decisions.

The preferences of the English public suggest that pain/discomfort and anxiety/depression are the health problems that are most important; while problems with self-care (ability to wash or dress oneself) and usual activities (e.g. ability to do work, study, housework, family or leisure activities) are less important. This reflects what members of the public deem important and has implications for the assessment of treatments that affect different aspects of HRQL.

This is one of the first studies internationally to report a value set for the EQ-5D-5L. A strength of the study is that the data have been generated using a standard protocol developed following an international programme of work (Oppe et al., 2014). Similar studies are now underway in numerous countries worldwide, which will facilitate direct comparisons of preferences between populations. See Augustovski et al. (in press), Ramos-Goñi et al. (in press), and Xie et al. (2016) for details of three studies that followed the same protocol. A further strength of our study is that it has addressed problems with previous value sets for EQ-5D, particularly with respect to the values for 'worse than dead' health states (Devlin et al., 2011; Devlin et al., 2012; Tilling et al., 2010), providing an improved basis for the use of HRQL evidence in decision making. Further, innovative methods developed during the course of our study have strengthened the approach to modelling value sets – in particular, by allowing different types of data (TTO and DCE) to be modelled together to providing complementary evidence on preferences; and by taking into account the nature of preference data that are 'bounded' (censored); heterogeneity of respondents' views in health utilities; and heteroskedasticity of the error terms.

Although our TTO data have good face validity, a potential limitation of our study is that there is evidence of clustering at certain values and of selective scale use. This could be linked to the relative difficulty of the TTO task, and the use of an automated process to guide its administration. Interviewer effects on TTO responses may also be important given potential differences in levels of interviewer abilities and engagement.

We therefore sought to exclude problematic data that could justifiably be considered not to reflect participants' true preferences, whilst avoiding the exclusion of inconvenient data based on subjective researcher judgements. This involved the exclusion of some participants' data and the censoring of some values at and below 0. The alternative approach of including all data would have meant that the assumption of a normal error distribution would need reconsidering. This in turn would have necessitated arbitrary assumptions, potentially with less transparency than the methods we report here. We deemed it invalid to include these data and to knowingly assume the wrong error distribution.

We tested an extensive range of model specifications in our econometric analysis, and each could feasibly have been used to generate a value set. The choice of the model reported in this paper necessarily reflects a number of researcher judgements about which model is 'best'. For example, while we could have generated a value set based on TTO data alone, the final value set reported here is derived from a hybrid of both TTO and DCE data, on the grounds that the two methods provide different and complementary information about the views of the sample.

Our choice of the least restrictive model is based on two considerations. First, there are strong prior grounds for selecting it, as it allows the values associated with different levels of problems to vary across the dimensions. Second, the overall model statistics, i.e. log likelihood, suggest that it better captures the nature of the preferences of the English general public than more restrictive models (Feng et al., 2015). Model fit was not improved by including interaction parameters, and so there are no such interaction terms in the value set. This is in contrast to the current EQ-5D value set, which included the so-called 'N3 term' (a parameter capturing an additional reduction in value to any health state with a level 3 problem on any dimension) (Dolan, 1997).

The EQ-5D-5L value set differs from the original EQ-5D value set (Dolan, 1997) and the interim crosswalk EQ-5D-5L tariff (van Hout et al., 2012) in important ways. First, the value for the worst state is higher – as expected, given well-known issues with the procedure for valuing health states worse than dead in the original value set study, which yielded values as low as -39 that required rescaling to -1. As well as a higher minimum value, the value set reported here also has considerably fewer states worse than dead (4.93%, compared to around a third in the original UK value set). The characteristics of the value set we report are more in line with those found in other countries (Szende et al., 2010). This may imply that the QALY gains for interventions targeting those in very poor health were overestimated previously.

The greater descriptive sensitivity of EQ-5D-5L allows patients to give more refined HRQL measurement data as they have more levels over which to describe their health. However, this increased ability to capture responses to treatment may be somewhat counteracted by the nature of the value set reported here, as members of the general public did not, on the whole, think that there was an important and significant difference between levels 4 and 5 on some dimensions. This highlights an interesting disjunction between the patient's perspective, when self-reporting their health on the EQ-5D-5L, and the perspective of the general public in valuing these states. Whose values should count in decision making using these data is a normative question (Brazier, 2008). Use of the general public's preferences is consistent with the stated requirements of NICE (2013) – but whether patients' preferences should be used instead is a question that should be given further consideration.

It would be possible to develop a range of value sets, based on the preferences of different population sub-groups and methodologies, for use in different contexts. Additional analyses (to be reported separately) show some differences between the health state preferences of different age groups, such that value sets estimated from age-specific data would differ in important ways. However, the use of a single value set, such as the one we report here, allows for consistent decision making across patient populations and sociodemographic groups – which is particularly important where resource allocation decisions are concerned.

This study raises a range of unanswered questions and areas for further research. First, the value set we have reported is for the English population. However, some health care decisions relate to different jurisdictions. For example, NICE decisions cover both England and Wales. The current UK EQ-5D value set is used by both NICE and the Scottish Medicines Consortium. Are the preferences of the UK population (i.e. including Scotland, Wales and Northern Ireland) consistent with the values reported here? We have collected additional data in order to generate a UK value set, which will be reported separately. Similarly, how do the English values compare with those produced in other countries? Over a dozen EQ-5D-5L value set studies are underway internationally, using the same protocol used in this study, and future research can compare these in detail.

Finally, although there is evidence to support the face validity of the data used to produce this value set, there are many remaining methodological issues that, if addressed, may help to further improve data quality. For example, changes to the way in which the stated preference tasks and health states are presented to participants may yield improved data. There are also a range of other promising preference elicitation methods that may be used to generate values, such as DCE designs that include an attribute for duration, and can therefore be modelled directly onto the 0 to 1 QALY scale (Mulhern et al., 2014; Bansback, et al., 2012). **Error! Reference source not found.**While new methods in this research area will continue to be developed, the value set reported here provides a robust and up-to-date basis for summarising EQ-5D-5L data in decision making.

REFERENCES

Augustovski, F., Rey-Ares, L., Irazola, V. and Garay, O.U., Gianneo, O., Fernández, G., Morales, M., Gibbons, L., Ramos-Goñi, J.M., In press. An EQ-5D-5L value set based on Uruguayan population preferences. *Qualty of Life Research*.

Bansback, N., Brazier, J., Tsuchiya, A. and Anis, A., 2012. Using a discrete choice experiment to estimate health state utility values. *Journal of Health Econ*omics, 31, pp.306-318.

Brazier, J., Ratcliffe, J., Salomon, J.A. and Tsuchiya, A., 2007. *Measuring and valuing health benefits for economic evaluation*. Oxford: Oxford University Press.

Brazier, J.,(2008) Valuing health states for use in cost-effectiveness analysis. *Pharmacoeconomics*, 26, pp.769-779.

Cambridgeshire Community Services NHS Trust. Quality Account 2013-2014.
http://www.nhs.uk/Services/Trusts/Overview/DefaultView.aspx?id=1118

Devlin, N. and Appleby, J., 2010. *Getting the most out of PROMs: putting health outcomes at the heart of NHS decision making.* London: Kings Fund.

Devlin, N., Buckingham, K., Shah, K., Tsuchiya, A., Tilling, C., Wilkinson, G. and van Hout, B., 2012. A comparison of alternative variants of the lead and lag time TTO. *Health Economics,* 22(5), pp.517-532

Devlin, N., Tsuchiya, A., Buckingham, K. and Tilling, C., 2011. A uniform Time Trade Off method for states better and worse than dead: feasibility study of the 'lead time' approach. *Health Economics,* 20(3), pp.348-361.

Dolan, P., 1997. Modeling valuations for EuroQol health states. *Medical Care,* 35(11), pp.1095-1108.

Feng, Y., Devlin, N.J., Shah, K., Mulhern, B. and van Hout, B., 2015. *New Methods for Modelling EQ-5D-5L value sets: an application to English data.* OHE Research Paper. London: Office of Health Economics.

Gold, M.R., Siegel, J.E., Russell, L.B. and Weinstein, M.C., 1996. *Cost-effectiveness in health and medicine.* New York: Oxford University Press.

Herdman, M., Gudex, C., Lloyd, A., Janssen, M.F., Kind, P., Parkin, D., Bonsel, G. and Badia, X., 2011. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qualife of Life Research*, 20(10), pp.1727-1736.

Janssen, B., Oppe, M., Versteegh, M. and Stolk, E., 2013. Introducing the composite TTO: a test of feasibility and face validity. *European Journal of Health Economics,* 14, pp.5-13.

Janssen, M.F., Lubetkin, E.I., Sekhobo, J.P. and Pickard, A.S., 2011. The use of the EQ-5D preference-based health status measure in adults with Type 2 diabetes mellitus. *Diabetes Medecine,* 28(4), pp.395-413.

Kind. P., Brooks, R. and Rabin, R., 2005. *EQ-5D concepts and methods: a developmental history.* Dordrecht: Springer.

Mulhern, B., Bansback, N., Brazier, J.E., Buckingham, K., Cairns, J., Dolan, P., Hole, A.R., Kavetsos, G., Longworth, L., Rowen, D. and Tsuchiya, A., 2014. Preparatory study for the revaluation of the EQ-5D tariff: methdology report. *Health Technology Assessment,* 18(12).

NICE, 2013. *Guide to the methods of technology appraisal 2013.* London: National Institute of Health and Care Excellence.

Office for National Statistics, 2011. *Census: Digitised Boundary Data* (England and Wales) [computer file]. UK Data Service Census Support. Available from: http://edina.ac.uk/ukborders

Oppe, M., Devlin, N. J., van Hout, B., Krabbe, P.F. and de Charro, F., 2014. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value in Health,* 17, pp.445-453.

Parkin, D., Devlin, N. and Feng, Y., 2014. *What determines the shape of an EQ-5D index distribution?* OHE Research Paper. London: Office of Health Economics.

Pickard, A.S., Wilke, C.T., Lin, H.W. and Lloyd, A., 2007. Health utilities using the EQ-5D in studies of cancer. *Pharmacoeconomics*, 25(5), pp.365-384.

Ramos-Goñi, J.M., Pinto-Prades, J.L., Cabasés, J.M. and Rivero-Arias, O., In press. Valuation and modeling of EQ-5D-5L health states using a hybrid approach. *Medical Care.*

Robinson, A. and Spencer, A., 2006. Exploring challenges to TTO utilities: valuing states worse than dead. *Health Economics,* 15, pp.393-402.

Ryan, M., Gerard, K., Amaya-Amaya, M., 2008 *Using discrete choice experiments to value health and health care*. Dordrecht: Springer.

Shah, K.K., Lloyd, A., Oppe, M. and Devlin, N., 2013. One-to-one versus group settings for conducting computer assisted TTO studies: findings from pilot studies in England and the Netherlands. *European Journal of Health Econ*omics, 14(Suppl 1), pp.S65-73.

Szende A., Oppe, M. and Devlin, N., 2007. *EQ-5D value sets: inventory, comparative review and user guide*. Dordrecht: Springer.

Tilling, C., Devlin, N., Tsuchiya, A. and Buckingham, K., 2010. TTO valuations of health states worse than dead: a literature review and conceptual framework for systematic analysis. *Medical Decision Mak*ing, 30, pp.610-619.

van Hout, B., Janssen, M.F., Feng, Y.S., Kohlmann, T., Busschbach, J., Golicki, D., Lloyd, A., Scalone, L., Kind, P., Pickard, A.S., 2012. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value in Health*, 15, pp.708-715.

Wailoo, A., David, S. and Tosh, J., 2010. *The incorporation of health benefits in CUA using EQ-5D*. Sheffield: NICE Decision Support Unit.

Xie, F., Pickard, A.S., Krabbe, P.F.M., Revicki, D., Viney, R., Devlin, N. and Feeny, D., 2015. A Checklist for Reporting Valuation Studies of Multi-Attribute Utility-Based Instruments (CREATE). *Pharmacoeconomics,* 33(8), pp.867-877.

Xie, F., Pullenayegum, E., Gaebel,K., Bansback, N., Bryan, S., Ohinmaa, A., Poissant, L., and Johnson, J.A., 2016. Canadian EQ-5D-5L Valuation Study Group. A Time Trade-off-derived Value Set of the EQ-5D-5L for Canada. *Medical Care*, 54(1), pp.98-105.