Developing an EQ-5D-5L value set for England

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

1. Office of Health Economics, London; 2. School of Health and Related Research, University of Sheffield; 3. Department of Economics, University of Sheffield

Acknowledgements
This study was partly funded by a Department of Health Policy Research Programme grant; the PRET data were funded by the MRC-NIHR Research Methodology Programme, and the PRET-AS data were funded by the EuroQol Group. Views expressed in the paper are those of the authors, and not the funding bodies. Research ethics approval has been given by the ethics committee at the School of Health and Related Research in line with University of Sheffield research governance and ethics requirements.

Abstract
Aims: The aim of this paper is to report the methods used to develop the EQ-5D-5L Value Set for England. The value set produced will be used by health care decision makers such as the English Department of Health and the National Institute for Health and Care Excellence (NICE). We address a number of methodological issues relating to the analysis of valuation data, including how conceptually different types of preference data – Time Trade Off (TTO) and Discrete Choice – can be combined in modelling health state values.

Methods: Respondents completed 10 composite TTO and seven DCE tasks. Values were elicited using face-to-face computer assisted personal interviews. Adult members of the general public in England were selected at random from postal addresses, and interviews were carried out by a team of 48 interviewers. Descriptive analyses were used to assess the quality of the data. A range of models were estimated for the TTO and DCE data, including 'hybrid' models estimated on pooled data, and also incorporating data from the 'Preparatory study for the re-valuation of the EQ-5D tariff' (PRET) project.

Results: Data were provided by 996 respondents; TTO data from 84 of these were excluded from modelling because they gave all states the same value, or gave 55555 a higher value than mild health states. 5, 9, 10 and 20-parameter models yielded coefficients with the expected magnitudes and signs. In all hybrid models, problems with pain/discomfort and anxiety/depression have the largest effect on utility. In the 20 parameter hybrid 'all' data model, values range from -0.16 for health state 55555 to 0.85 for health state 21111. Models which account for heterogeneity in preferences yield more extreme values, e.g. a value of -0.3 for 55555.

Discussion: Respondents are more concerned about problems with pain/discomfort and anxiety/depression than problems with mobility, self-care and usual activities. Compared to the existing UK value set for EQ-5D, relatively few health states were valued as being worse than dead, and the minimum estimated value (for health state 55555) is relatively high. We welcome feedback from HESG participants on a number of remaining methodological issues.

Introduction

The EQ-5D (Brooks, 1996) is a 'generic' measure of patient-reported health with an accompanying population value set (Dolan, 1997) that is widely used in England and internationally. In addition to its use in academic studies in the UK, it is the instrument recommended by the National Institute for Health and Care Excellence (NICE) for evidence submitted to its health technology assessment (HTA) process (NICE 2013); is included in population health surveys such as the Health Survey for England; and is used in the English NHS PROMs programme (Devlin and Appleby 2010).

While there is evidence that the EQ-5D is a valid and reliable measure of health-related quality of life in many disease areas (Pickard et al 2007; Janssen et al 2011; Wailoo et al 2010), there have been concerns about the possibility of ceiling effects limiting its ability adequately to capture milder health problems (i.e. in the step from no problems to moderate problems); and therefore its sensitivity to small changes in health.

In response to these concerns, the EuroQol Group developed a new version of the instrument, the EQ-5D-5L (Herdman et al 2011), aimed at increasing the sensitivity of the instrument. The EQ-5D-5L comprises the same core five dimensions (mobility; self-care; usual activities; pain/discomfort; anxiety/depression) included in the EQ-5D, but increases the available response options from three (no, some, or extreme problems) to five (no, slight, moderate, severe, or extreme problems/unable to). The five dimensions and five levels of the EQ-5D-5L describe (5^5 =) 3,125 health states, compared to the (3^5 =) 243 described by the EQ-5D.

The EQ-5D is accompanied by 'value sets' which provide, for each health state described by the instrument, a number summarising how good or bad that health state is, on a scale anchored at 1 (full health) and 0 (a state as bad as being dead), with numbers < 0 representing health states considered worse than being dead. These value sets, which are available for a range of countries (Szende et al 2007), are required for the use of patients' EQ-5D data in the estimation of quality adjusted life years (QALYs) (Drummond et al 2005). The value sets are also used to summarise patients' self-reported EQ-5D data for the purposes of statistical analysis (Parkin et al 2010), for example, in the NHS PROMs programme (HSCIC 2013). The 'values' reported in these value sets are based on the views of a sample of the general public, asked to imagine living in health states that are hypothetical to them and to respond to a series of structured questions designed to discover their preferences about the different aspects of health described by the EQ-5D. This approach follows the requirements of NICE and similar HTA organisations, 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 those of patients (Gold et al 1996).

Research has been undertaken to establish the relationship between patients' self-reported health on the EQ-5D and on the EQ-5D-5L. This has produced an algorithm to enable EQ-5D-5L data to be 'mapped' to EQ-5D health states, and to the corresponding value sets available for each country (van Hout et al 2012). Whilst this provides an interim way of using EQ-5D-5L data, such an approach is subject to a number of important limitations - including the limitations of the UK EQ-5D value set itself. It is widely acknowledged that there were problems with the way observations with negative values were handled in that study, and a general view is that too many of the health states in the value set (approximately a third) have negative modelled values (i.e. are worse than dead). This has important implications for its application in estimating QALY gains (Devlin et al 2011; 2013).

Ultimately, a value set based directly on preferences of the general public for EQ-5D-5L health states is required. To date, there are no such value sets available for the EQ-5D-5L. The aim of this study is to test methods that can be used to produce a value set for the EQ-5D-5L, based on the preferences of the general public in England. The subsequent value set can be used to support decision making in the English NHS.

Methods

Methods of eliciting preferences

The study used the EuroQol Group Valuation Technology (EQ-VT), developed specifically for EQ-5D-5L value set studies, administered using computer-assisted personal interviews (CAPI) in the homes of respondents. The methods selected to elicit preferences were an adapted version of the Time Trade Off (TTO) – the approach used in previous valuation studies of EQ-5D, and accepted by NICE as a 'choice-based' approach – supplemented with Discrete Choice Experiment (DCE) questions, an approach increasingly used to assess preferences for health states (Ryan et al 2008).

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); background questions regarding age, gender and experience of serious illness; a practice TTO task to introduce the process of valuing states both better and worse than dead (using a health state describing wheelchair use); 10 TTO tasks; three structured feedback questions regarding the TTO tasks; seven DCE tasks; three structured feedback questions regarding the DCE tasks; an opportunity to leave feedback via an open-ended comment box; and further background questions. After each interview, the interviewer completed a separate questionnaire comprising three questions about the concentration level and understanding of the respondent, and the environment in which the interview was conducted.

In the TTO tasks, a composite approach was used. This involved beginning with the 'conventional' TTO for all health states, and shifting to a 'lead time' TTO (Robinson and Spencer 2006; Devlin et al 2011 and 2013) if respondents indicated that they considered the health state to be worse than being dead. Evidence on the composite TTO approach with EQ-VT (illustrated in Figures 1a and 1b) is reported by Janssen et al (2013).

Figure 1a illustrates the TTO task for states better than dead (i.e. have a value > 0). The respondent is asked to imagine living for 10 years in the EQ-5D-5L state described in the blue box ('Life B'), followed by death. The value of that health state is then identified by finding, using an iterative process, the number of years in full health ('Life A') they consider equivalent to that. The iterative process for a state better than dead starts with Life A at 10 years (equivalent to a value of 1), moves to 0 years (equivalent to a value of 0), then 5 years (0.5), then either 4 (0.4) or 6 (0.6) years (and so on) based on the respondent's choice. Half year values are also available to respondents depending on their previous choice.

The more severe the health state, the more years of full health the respondent will generally be willing to give up in Life A. For very poor health states, all the time in Life A may be traded off – indicating that the value for the state is zero or less (i.e. is equal to or worse than immediate death). Where this occurs, additional time in full health is added to the start of both Life A and Life B – see 1b – allowing respondents to trade off more time to reflect that the extent to which they consider the health state to be worse than dead. The variant of lead time TTO used in this study involves a 20 year time frame (10 years of lead time followed by 10 years in the health state). The iterative process for negative values starts with Life A at 10 years (equivalent to a value of 0), then moves to 5 years (-0.5) and then 4 years (-0.6) or 6 years (-0.4). The minimum value that can be produced directly from this variant is -1, where all of the time in Life A has been traded. No additional trade-off questions were asked of those who 'exhausted' their lead time, so respondents whose actual value is lower than -1 are not identified.

The automated iterative routing used to seek the point of indifference was based on an adaptation of the Dolan (1997) EQ-5D value set study. Further details of the EQ-VT and the iterative process used in the TTO tasks are available in Oppe et al (2013).

Each TTO task ends when the respondent clicks the button indicating that they consider Life A and Life B to be 'about the same'. At this point of indifference, the implied value for 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 U=t/10, where U is health state value (utility) and t is the number of years in Life A at the respondent's point of indifference. The implied value for 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 U=(t-10)/10.

The maximum value of 1 implies that the respondent considers the health state to be as good as being in full health. A value of zero implies that the respondent considers the health state to be no better and no worse than being dead. The value of -1 implies that the respondent considers the health state to be so undesirable that they are indifferent between dying immediately and living for 10 years in full health followed by 10 years in the health state. However, as the minimum possible value was -1, values at -1 are effectively censored data.

In each DCE task (Figure 2), respondents were presented with a pair of health states (A and B), with no reference to the duration of the states, and were asked to indicate which they considered to be 'better'. No indifference option was included.

Study design

The EuroQol Group's research protocol for EQ-5D-5L value set studies employs a blocked design to select 86 health states (from the 3,125 defined by the EQ-5D-5L) for TTO valuation and 200 pairs of states for valuation via DCE (Oppe et al 2013). Respondents were randomly assigned to one of 10 blocks of TTO tasks and to one of 28 blocks of DCE tasks. The order in which the states appeared in interviews was also randomised. None of the pairs in the DCE design included one health state that could be said to 'logically dominate' the other. Each block of TTO tasks contained a combination of mild, moderate and severe health states. All TTO blocks included the worst health state (55555) and one of the mildest possible health states (with no problems on four dimensions and slight problems on one dimension, for example 21111).

Data collection

The fieldwork for this study, including sample recruitment and interviewing, was carried out by Ipsos MORI. A sample of 2,020 addresses from 66 primary sampling units (based on postcode sectors) across England was randomly selected, using the Post Office small user Postcode Address File as the sampling frame. 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. Interviewers working for Ipsos MORI sent an advance letter and information sheet to each selected address, together with a small, unconditional incentive of six first class stamps. For each address, interviewers were required to make a minimum of six personal calls at different times and on different days of the week. In cases where a given address generated more than one dwelling unit, one of the dwelling units was selected randomly using a selection grid. In each selected dwelling unit, all individuals aged 18 years and over were listed in alphabetical order of first name and one was selected randomly using a selection grid (no substitutes were permitted).

If the selected individual consented to take part in the study they were interviewed in their own home. The respondent was in control of the laptop throughout the tasks (but not the practice task), with the interviewer guiding them through each step, following a script. The one-to-one setting allowed interviewers to provide detailed instruction and feedback to the respondents where appropriate.

The majority of the interviewers were experienced in carrying out health-related surveys with members of the general public, but few had previously conducted interviews involving TTO. In total, 48 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.

Throughout the data collection period, the interim data were monitored by the research team on a regular basis (at least weekly; some statistics were monitored on a daily basis during the first few months of data collection). All data were monitored at the interviewer level; if a given interviewer was found to be generating unusual or poor quality data (defined using a number of criteria including the frequency of 'round number' values 1, 0.5 and 0; and the amount of time taken to complete the tasks – see Mulhern et al (2013) for details), that interviewer was contacted by Ipsos MORI and given additional training. Further advice and instruction was provided to interviewers via Ipsos MORI's periodic study 'newsletter'.

The Preparatory Study for the Re-valuation of the EQ-5D Tariff (PRET/PRET-AS) study

To inform the development of the EQ-5D-5L vale set, selected data from the PRET/PRET-AS study (Mulhern et al., in press) were also used. PRET was a methodological study investigating a range of issues relating to TTO and DCE using online methods and face-to-face interviews. The data used were taken from a large online DCE study incorporating an attribute for duration (DCE_{TTO}), a method which has been shown to be feasible for valuing EQ-5D (Bansback et al. 2012; Viney et al. 2013) and EQ-5D-5L (Norman et al. 2013) health states. In the PRET-AS DCE_{TTO} study, 120 pairs of EQ-5D-5L health states were paired with one of three levels of duration (1 year, 5 years, 10 years) and administered to 1,800 members of an online panel representative of the UK population in terms of age and gender. Only 18 pairs included differences in duration and gave (implicit) information about whether a health state is better or worse than dead.

Methods of analysis

Detailed descriptive analyses were undertaken of the quantitative and qualitative data provided by respondents – for a full report, see Mulhern et al (2013). Various exclusion rules were considered, following scrutiny of the scatterplots of all 996 respondents' individual-level regressions between TTO values and health state 'misery' scores (a proxy for severity, defined as the sum of the EQ-5D-5L levels for the health state valued; e.g. the misery score for state 22213 = 2+2+2+1+3 = 10) by three members of the research team. This provided a systematic way of identifying problem 'types'. We developed exclusion criteria with the aim of excluding from the modelling TTO data that very clearly suggested a lack of understanding or engagement, whilst minimising heroic researcher judgements and also taking into account the exclusion criteria used in other similar studies (Szende et al. 2007). The criteria for exclusions were: (a) all TTO health states given the same value; and (b) health state 55555 given a higher value than the mildest health state in the block. Further, there were a number of cases where 55555 was valued at 0 but logically better states were given a negative value. Rather than discarding these data, negative values in these cases were set to zero and then treated as being censored at zero.

Data modelling

Models with a variety of alternative specifications of the parameters were tested, each providing different ways of capturing observed differences in preferences between dimensions and levels; together with a variety of methods for testing for possible interaction effects between dimensions. Given the substantial heterogeneity commonly observed in health state preference data, our modelling strategy included alternative ways of handling this, including random coefficient models, which estimate value functions for every individual member of the sample.

The main specifications included a 5 parameter model (one parameter for each dimension), a 9 parameter model (one parameter for each dimension and four parameters reflecting four severity levels: slight, moderate, severe, and unable/extreme), a 10 parameter model (one parameter for each dimension and five parameters reflecting five severity levels) and a 20 parameter model (four parameters for each of the five dimensions reflecting a utility decrement for each severity level). All models were estimated for both TTO and DCE (and DCE_{TTO}) data, and 'hybrids' of the these.

The estimations from TTO and DCE data generate different parameters. The hybrid approach described in Rowen et al (2011) and Oppe and van Hout (2010) is used to bring both estimates together using the likelihood functions from both models. The approach takes a maximum likelihood estimator approach. To illustrate this, in the standard 5 parameter model for the TTO data we have:

$$v_{ij} = \alpha + \beta' d_{ij} + e_{ij}$$
$$e_{ij} \sim N(0, \sigma)$$

where v_{ij} is the TTO value that individual i scored for health state j. d_{ij} is a vector that includes the 5 dimensions of a health state j. e_{ij} is the error term which assumed to be normally distributed with zero mean and standard error σ. v_{ij} can be censored at either 0 or -1, i.e. $\alpha + \beta' d_{ij} + e_{ij} \leq 0$ or $\alpha + \beta' d_{ij} + e_{ij} \leq -1$.

The log likelihood function from the TTO data is shown as below:

$$\text{Loglik(tto)} = \log\left(\prod_{1}^{7579} f(e_{ij})\right) + \log\left(\prod_{1}^{1187} P(\alpha + \beta' d_{ij} < 0)\right) + \log\left(\prod_{1}^{354} P(\alpha + \beta' d_{ij} < -1)\right)$$

When considering the DCE data, respondents compare the values of two health states, i.e. v_{ijl} and v_{ijr}. We formulise the comparison as follows:

$$v_{ijl} = \alpha + \beta' d_{ijl} + e_{ijl} <?> v_{ijr} = \alpha + \beta' d_{ijr} + e_{ijr}$$

When assuming a logistic[1] distribution for the difference in the errors, the likelihood function is

$$\text{Loglik(dce)} = \log\left(\prod_{1}^{3642} \frac{\exp\left(\beta'(d_{ijl} - d_{ijr})\right)}{1 + \exp\left(\beta'(d_{ijl} - d_{ijr})\right)} + \log\left(\prod_{1}^{3330} \left(1 - \frac{\exp\left(\beta'(d_{ijl} - d_{ijr})\right)}{1 + \exp\left(\beta'(d_{ijl} - d_{ijr})\right)}\right)\right)\right)$$

The hybrid method assumes that the relative weights resulting from the TTO and DCE data are the same:

$$\hat{\beta}(\text{tto})'\theta + \gamma = \hat{\beta}(\text{dce})'$$

The joint log likelihood function from the TTO and DCE data is shown as below:

$$\begin{aligned} &\text{Loglik(tto and dce)} \\ &\quad = \log\left(\prod_{1}^{7579} f(e_{ij})\right) + \log\left(\prod_{1}^{1187} P(\alpha + \beta' d_{ij} < 0)\right) + \log\left(\prod_{1}^{354} P(\alpha + \beta' d_{ij} < -1)\right) \\ &\quad + \log\left(\prod_{1}^{3642} \frac{\exp\left((\beta'\theta + \gamma)(d_{ijl} - d_{ijr})\right)}{1 + \exp\left((\beta'\theta + \gamma)(d_{ijl} - d_{ijr})\right)}\right) \\ &\quad + \log\left(\prod_{1}^{3330} \left(1 - \frac{\exp\left((\beta'\theta + \gamma)(d_{ijl} - d_{ijr})\right)}{1 + \exp\left((\beta'\theta + \gamma)(d_{ijl} - d_{ijr})\right)}\right)\right) \end{aligned}$$

A similar approach was used for the inclusion of the likelihood of the DCE_{TTO} data as was used with the DCE data. Now time enters the comparison as follows:

$$v_{ijl} = t_{ijl}(\alpha + \beta' d_{ijl}) + e_{ijl} <?> v_{ijr} = t_{ijr}(\alpha + \beta' d_{ijr}) + e_{ijr}$$

And a separate linear transformation was used to reflect the assumption that the relative weights between the coefficients in TTO and DCE_{TTO} are the same.

[1] A logistic model was chosen instead of Probit model because of the expectation of no difference in results, but the clear advantage in computing time.

All models are estimated by maximizing the likelihood as well as by using WinBugs, using uninformative priors on all parameters.

As noted earlier, the minimum value possible given the design of the composite TTO task is -1. While this avoids the need for arbitrary *post hoc* rescaling to -1 (as in Dolan 2007 and similar studies), it means that the TTO data are in effect censored at -1, since *some* of these individuals may have wished to assign values < -1 for very poor states (see Devlin et al 2012) had that been possible.. Heterogeneity in the TTO values is modelled by a random coefficient model with an explicit parameter which defines the slope of each individual from the best to the worst health state. Two specifications for the distribution of the slope of the respondents are tested. First the slope is assumed to either have a lognormal distribution with mean 1. Second a nonparametric distribution is assumed with density on 3 different points, assuming that there are 3 types of individuals and that people belong to either one of those three groups. Statistical analysis and modelling were performed in STATA, Winbugs and R.

Results

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 and DCE tasks, and basic background questions) in full. The mean time taken to complete all the valuation tasks was 21.0 minutes (range: 1.3 to 114.3 minutes). The response rate was approximately 40%. Full sociodemographic and wellbeing data were collected for 985 of the 996 respondents (98.9%). The background characteristics of the sample, and of the sample whose data are available from the PRET study, are presented in Table 1. 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 who self-report health problems.

Descriptive analysis

Figure 3 shows the distribution of all TTO responses. The majority of values are positive with few values between 0 and -0.5, and clusters of values at 1, 0.5 and 0. There were a substantial number of inconsistent responses (for example valuing the worst state, 55555, higher than other states). Interviewer effects were apparent, with some interviewers more likely to record values across the full range of the scale, or at certain points of the scale, than others. Figure 4 shows that the mean and median TTO values by health state decrease as the severity of the state, measured by its misery score, increases. The mean TTO of the states with misery score of 6 is 0.86 (median = 0.95). The mean value of 55555 (completed by all respondents) is -0.01 (median = 0.00). The mean number of steps taken to complete a TTO task was 6.2, with more than 90% of tasks completed in less than 10 steps. Many final values were reached using the minimum number of steps required for that value. For example, of all of the valuations of 0.5, 86% of these were reached using three steps. For DCE, Figure 5 displays the relationship between the proportion of respondents choosing A or B and the relative severities of those states (again, measured by their misery scores), and shows that the majority of respondents consistently chose the less severe health state. When states were similar, there was little difference between the proportions of respondents choosing A or B. A detailed descriptive analysis of the TTO and DCE data is reported in Mulhern et al (2013).

Exclusion criteria and developing the modelling dataset

We excluded 23 respondents who gave all 10 health states the same value; and 61 respondents who valued 55555 (misery score = 25) no lower than the value they gave to the mildest health state included in their block (misery score = 6). The core modelling dataset therefore includes 912 respondents, with 10 TTO observations for each. Of the remaining respondents, 65 had valued 55555 at 0 but given negative values to other states which, in the modelling data set, were adjusted to zero, and those zero values treated as censored data.

Model results

The Deviance Information Criterion (DIC) suggests the 20 parameter model outperforms the 5, 9 and 10-parameter models. The main results from the 20 parameter model using the hybrid approach are reported in Table 2; Appendix 1 reports the results of the 5, 9 and 10 parameter models for comparison. The coefficients for levels 4 and 5 for depression are wrongly ordered in the linear model of TTO data, but correct in the DCE-models and the hybrid models. This is also true when considering the 10 parameter models where the relative levels of severe and extreme are assumed to be identical for pain/discomfort and anxiety/depression.

All models and all data-sets show that mobility, pain/discomfort and anxiety/depression get the highest weight. Analysis and detailed assessment of the extent to which the models and datasets differ is ongoing.

The results in Table 3 suggest that the predicted value of health state 55555 is -0.153 when combining the DCE and TTO data, and -0.161 when combining the DCE, TTO and DCE_{TTO} PRET-AS data. Examples of the how the results from the 20-parameter model are used in the calculation of values for EQ-5D-5L health states are shown in Table 3.

As expected, the hybrid models strike a compromise between the DCE and TTO models.
The results from the random coefficients models are reported in Appendix 2. The results suggest that the predicted value of 55555 decreased to -0.197 using the random coefficient model, -0.358 from the random coefficient model with random slope on 3 points, and -0.387 from random coefficient model with lognormal random slope.

Discussion

We have reported the valuation and modelling methods used to develop a value set for EQ-5D-5L for use in health care decision making in England. Given the importance of this work to the health care decision making processes, our analyses paid particular attention to TTO data quality issues. There were relatively high levels of logically inconsistent responses to the TTO tasks and, more generally, a high frequency of responses centred on a handful of 'round number' values (e.g. 0, 0.5, 1). Prior to exclusions, models estimated on the full TTO dataset yielded inconsistent coefficients on some dimensions. These issues may reflect the greater complexity for respondents in imagining, differentiating between, and valuing health states described in terms of five levels rather than three (Shah et al 2013). The composite TTO methodology uses two different methods to elicit values for states better than or worse than dead, and this may be complex for some respondents. The use of

CAPI to administer the tasks in the current study, which was not a feature of the Dolan (1997) value set study, may have reduced the quality of the interaction between interviewers and respondents.

The modelling approaches we have used rely on a number of assumptions. First, values of -1 are interpreted as possibly indicating values < -1. This may not be the case where respondents exhaust their lead time because they misunderstood the task, or are simply signalling a qualitative view that the health state is very bad without being able to specify an actual value (Devlin et al 2012). Second, some values of 0 are also treated as censored, where they came from respondents who did not value 55555 as the worst state. These respondents are assumed to be making a mistake of some kind beyond the scale of measurement errors as we usually understand them.

We elicited TTO values for EQ-5D-5L health states relative to 'full health' (see Figure 1) rather than to best health described by EQ-5D-5L (11111), as in Dolan et al (1997). Arguably, this is more in keeping with the way utilities are used in cost effectiveness analysis, where 1 = full health (and given the evidence that people do not consider 11111 to constitute full health). The constant term in the hybrid models might be interpreted as any departure from full health and, in the absence of problems on any dimension, might be thought to represent the value of 11111. The value of 11111, and how (conceptually and empirically) to interpret the gap between 11111 and full health has never previously been investigated. We are currently engaged in further research to address these research questions.

We have reported a number of models in this paper. The appropriate criteria for choosing between them is not obvious; yet the implications of that choice are potentially very important for their use in decision making. For example, the 20 parameter 'all' data hybrid value set reported in this paper has a minimum value of -0.16 for the worst health state (55555), and a maximum value of 0.86 for health state 21111. This is smaller than the range of values in the current EQ-5D value set (Dolan 1997), where the minimum value (for 33333) was -0.594; and the highest value (for 11211) was 0.883. The low minimum value in the EQ-5D value set reflects a methodological issue with that study; improved TTO methods for handling values < 0 used in this study have, as we expected, led to higher values for very bad health states. However, we also expected mild states in the EQ-5D-5L to have higher values than mild states in the EQ-5D, as the presence of 'some/moderate' problems in the EQ-5D is arguably worse than 'slight' problems on the EQ-5D-5L.

The hybrid modelling approach has been central to our modelling approach. This approach assumes that when people answer TTO questions that they use as similar underlying concept of value as when they answer DCE (or DCE_{TTO}) questions. Whether this is a heroic assumption is subject of further research.

The smaller range of values in value set reported here will have implications for estimates of improvements in health using EQ-5D-5L data. While EQ-5D-5L data collected from patients will have greater descriptive sensitivity than EQ-5D data the valuation of those states and improvements in them will not necessarily lead to estimates of greater QALY gains. It is possible that patients themselves would value these health states differently from the general public, which might suggest revisiting either the normative principle or the practice of seeking values from the general public rather than patients (Stamuli 2011).

The focus of this study was to produce an EQ-5D-5L value set for England. However, given that some decision making processes using patient-reported outcomes data relate to the UK more generally (such as the proposed Value Based Pricing scheme) we have undertaken additional data collection in Scotland, Northern Ireland and Wales which, combined with the English data reported here, will provide the basis for a UK value set. That work will be reported separately.

We welcome feedback on any aspect of this paper, but especially welcome views on the following:

1. Is the way we have handled data quality issues (eg. the exclusion criteria used) defensible?
2. Our advisory group, including representatives from NICE and the DH, has indicated a strong preference for us to report one 'main' value set from this work, rather than publishing a suite of value sets. What criteria should we use to choose between our models?
3. Should we assume that health state 11111 has a value of 1, even though we know that people self-reporting their health state as 11111 on EQ-5D (3L and 5L) do not consider themselves to be 'full health'?
4. At present, the error distribution for the value of 1 is symmetric, which is likely to be wrong and is likely to lead to an underestimate of the true value of the good health states. How should this be handled?

References

Brooks, R. (1996) EuroQol: The current state of play. *Health Policy.* 37: 53-72.
Devlin, N. and Buckingham, K. (2013) *What is the normative basis for selecting the measure of 'average' preferences to use in social choices?* Proceedings of the Scientific Plenary Meeting of the EuroQol Group, Montreal.
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): 517-532.
Devlin, N.J., 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): 348-361.
Devlin, N., Hansen, P., Kind, P. and Williams, A. (2003) Logical inconsistencies in survey respondents' health state valuations – a methodological challenge for estimating social tariffs. *Health Economics* 12(7): 529-544.
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.
Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care* 35(11): 1095-1108.
Drummond, M.F., Sculpher, M.J., Torrance, G.W., O'Brien, J. and Stoddart, G.L. (2005) *Methods for the economic evaluation of health care programs*. Oxford: Oxford University Press.
Feng, Y., Devlin, N., Shah, K., Mulhern, B., van Hout, B. (2013) *The value of zero and the value of minus 1.* Proceedings of the Scientific Plenary meeting of the EuroQol Group, Montreal.
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.
HSCIC (Health and Social Care Information Centre). (2013) *Provisional monthly patient reported outcome measures (PROMs) in England. A guide to PROMs methodology*. London: HSCIC. Available at: http://www.hscic.gov.uk/media/1537/A-Guide-to-PROMs-Methodology/pdf/PROMS_Guide_v5.pdf

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). *Quality of Life Research* 20(10): 1727-1736.
Janssen, B., Oppe, M., Versteegh, M., Stolk, E. (2013) Introducing the composite TTO: a test of feasibility and face validity. *European Journal of Health Economics.* 14(1): 5-13.
Mulhern, B., Bansback, N., Brazier, J.E., Buckingham, K., Cairns, J., Dolan, P., Hole, A.R., Kavetsos, G., Longworth, L., Rowen, D., Tsuchiya, A. (in press) Preparatory Study for the Re-valuation of the EQ-5D tariff: Project report. *Health Technology Assessment*.
Mulhern, B., Shah, K.K., Feng, Y., Devlin, N., van Hout, B. (2013) *Assessing EQ-5D-5L valuation study data quality.* Proceedings of the EuroQol Scientific Plenary Meeting, Montreal.
NICE (2013) Guide to the methods of technology appraisal 2013. London: National Institute for Health and Care Excellence.
Office for National Statistics (2011) *Census: Digitised Boundary Data (England and Wales) [computer file]. UK Data Service Census Support*. Downloaded from: http://edina.ac.uk/ukborders
Oppe, M., Devlin, N.J, van Hout, B., Krabbe, P. and de Charro, F. (2013) A programme of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value in Health.* (forthcoming)
Oppe, M. and van Hou,t B. (2010) *The optimal hybrid: experimental design and modelling of a combine of TTO and DCE.* Proceedings of the Scientific Plenary Meeting of the EuroQol Group, Athens.
Rowen D., Brazier J. and van Hout, B. (2011) *A comparison of methods for converting DCE values onto the full health-dead QALY scale*. HEDS Discussion Paper 11/15. Sheffield: ScHARR.
Ryan, M., Gerard, K. and 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 Economics* 14(S1): 65-73.
Stamuli, E. (2011) Health outcomes in economic evaluation: who should value health? *British Medical Bulletin.* 97(1): 197-210
Szende, A., Oppe, M. and Devlin, N. (2007) *EQ-5D valuation sets: an inventory, comparative review and users' guide.* Dordrecht: Springer.
Parkin, D., Rice, N. and Devlin, N. (2010). Statistical analysis of EQ-5D profiles: does the use of value sets bias inference? *Medical Decision Making.* 30(5): 556-565.
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): 365-84.
van Hout, B., Janssen, M.F., Feng, Y-S., Kohlmann, T., Busschbach, J., Golicki, D., Lloyd, A., Scalone, L., Kind, P. and 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(5): 708-15.
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. *Diabetic Medicine* 28(4): 395-413.
Wailoo, A., David, S. and Tosh, J. (2010) *The incorporation of health benefits in CUA using EQ-5D*. Sheffield: NICE Decision Support Unit.

Table 1. Background characteristics of the sample

	Value set All respondents		Value set After exclusions		PRET-AS		General population [i]
	#	**%**	**#**	**%**	**#**	**%**	**%**
Age							
18-29	113	11.3%	105	11.5%	452	25.1%	20.7%
30-44	298	29.9%	270	29.6%	596	33.1%	26.3%
45-59	250	25.1%	227	24.9%	569	31.6%	24.7%
60-74	207	20.8%	191	20.9%	176	9.8%	18.5%
75+	128	12.9%	119	13.0%	0	0.0%	9.9%
Gender							
Male	405	40.7%	372	40.8%	820	46.6%	49.2%
Female	591	59.3%	540	59.2%	979	54.4%	50.8%
Economic activity							
Employed or self-employed	504	51.2%	463	50.8%	1042	57.9%	59.4%
Retired	278	28.2%	256	28.1%	110	6.1%	13.1%
Student	20	2.0%	19	2.1%	153	8.5%	8.8%
Looking after home or family	83	8.4%	73	8.0%	169	9.4%	4.2%
Long-term sick or disabled	48	4.9%	42	4.6%	136	7.6%	3.9%
Other / none of the above	52	5.3%	47	5.2%	189	10.5%	10.6%
Marital status							
Never Married	238	24.2%	0	23.4%	588	32.9%	34.6%
Married	466	47.3%	434	47.6%	1019	56.6%	46.6%
Same-sex civil partnership	2	0.2%	2	0.2%	N/A	N/A	0.2%
Separated [ii]	37	3.8%	32	3.5%	42	2.3%	2.7%
Divorced	131	13.3%	119	13.0%	118	6.6%	9.0%
Widowed [iii]	107	10.9%	99	10.9%	24	1.3%	6.9%
Prefer not to say	4	0.4%	1	0.1%	8	0.4%	N/A
Religion							
Christian	636	64.6%	575	63.9%	N/A	N/A	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%	N/A	N/A	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%	N/A	N/A	5.6% [iv]
Limited a little	158	16.0%	144	16.0%			7.1% [iv]
Not limited	716	72.7%	661	73.4%			87.3% [iv]
Education							
Degree	211	21.2%	201	22.3%	762	42%	
No degree	785	78.8%	699	77.7%	1037	58%	
Main language spoken							
English	920	93.4%	847	94.1%	N/A	N/A	
Any other language	65	6.6%	53	5.9%			
Responsibility for children							
Yes	350	35.5%	314	34.9%	N/A	N/A	
No	635	64.5%	586	65.1%			
Experience of serious illness							
In self	330	33.1%	297	33.0%	N/A	N/A	
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	48.6%	650	36.1%	
Any other health state	522	52.4%	475	52.8%	1149	63.9%	
Self-rated health using EQ-VAS							
<80	334	33.5%	298	33.1%	N/A	N/A	
80-89	256	25.7%	241	26.8%			
90-99	337	33.8%	306	34.0%			
100	69	6.9%	67	7.4%			

[i] General population data based on results of the 2011 UK Census (ONS 2011), where available

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

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

[iv] Census data reported here refer to individuals aged 16-64 only

Table 2. Results from the 20 parameter models for TTO, DCE, TTO and DCE hybrid; DCE (PRET), and 'all data' hybrid.

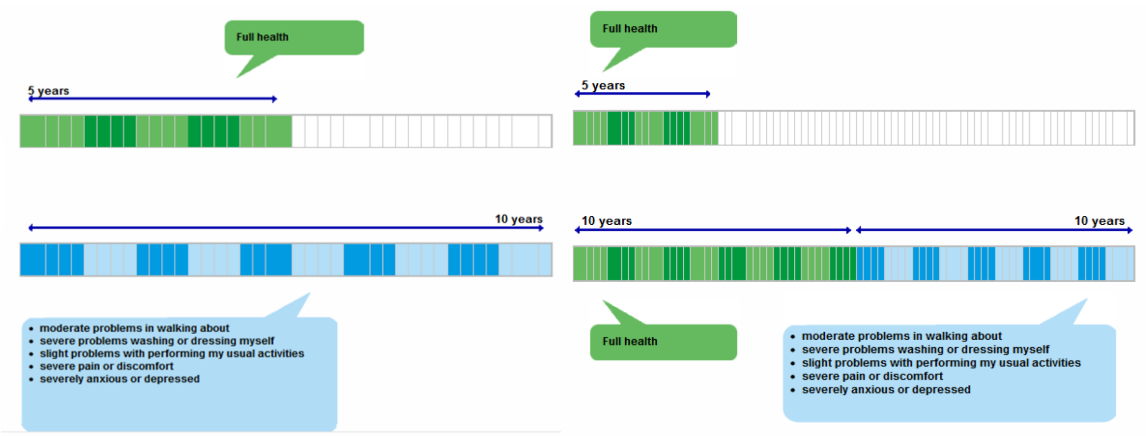
	tto linear model	dce dce logistic model	dce with duation logistic model	tto + dce hybrid model	all hybri model
constant	0.886 (0.854 - 0.919)			0.894 (0.870 - 0.918)	0.877 (0.857 - 0.896)
MO=2	-0.024 -(0.050 - 0.004)	0.347 (0.234 - 0.461)	-0.006 -(0.015 - 0.003)	-0.050 -(0.065 - -0.035)	-0.022 -(0.032 - -0.012)
MO=3	-0.046 -(0.073 - -0.019)	0.439 (0.309 - 0.570)	0.022 (0.014 - 0.030)	-0.066 -(0.084 - -0.048)	-0.044 -(0.054 - -0.034)
MO=4	-0.158 -(0.195 - -0.119)	1.126 (1.002 - 1.264)	0.096 (0.087 - 0.107)	-0.179 -(0.197 - -0.162)	-0.153 -(0.165 - -0.142)
MO=5	-0.209 -(0.245 - -0.173)	1.433 (1.293 - 1.583)	0.115 (0.108 - 0.124)	-0.229 -(0.247 - -0.212)	-0.188 -(0.199 - -0.177)
SC=2	-0.031 -(0.059 - -0.003)	0.267 (0.144 - 0.387)	0.003 -(0.006 - 0.012)	-0.043 -(0.060 - -0.027)	-0.023 -(0.033 - -0.013)
SC=3	-0.049 -(0.085 - -0.014)	0.404 (0.274 - 0.536)	0.022 (0.014 - 0.031)	-0.065 -(0.083 - -0.048)	-0.044 -(0.055 - -0.034)
SC=4	-0.095 -(0.133 - -0.057)	1.003 (0.870 - 1.140)	0.102 (0.094 - 0.110)	-0.150 -(0.169 - -0.133)	-0.147 -(0.158 - -0.135)
SC=5	-0.177 -(0.211 - -0.140)	1.046 (0.919 - 1.177)	0.133 (0.123 - 0.143)	-0.175 -(0.192 - -0.158)	-0.183 -(0.194 - -0.171)
UA=2	-0.048 -(0.082 - -0.014)	0.211 (0.104 - 0.323)	0.024 (0.014 - 0.033)	-0.038 -(0.054 - -0.022)	-0.038 -(0.049 - -0.027)
UA=3	-0.076 -(0.113 - -0.036)	0.213 (0.094 - 0.331)	0.048 (0.040 - 0.056)	-0.042 -(0.060 - -0.025)	-0.060 -(0.070 - -0.048)
UA=4	-0.127 -(0.164 - -0.093)	0.794 (0.670 - 0.924)	0.082 (0.075 - 0.092)	-0.131 -(0.149 - -0.113)	-0.128 -(0.140 - -0.117)
UA=5	-0.154 -(0.189 - -0.119)	0.811 (0.684 - 0.941)	0.092 (0.085 - 0.100)	-0.139 -(0.156 - -0.121)	-0.135 -(0.146 - -0.124)
PD=2	-0.033 -(0.064 - -0.001)	0.328 (0.215 - 0.446)	0.028 (0.020 - 0.038)	-0.047 -(0.064 - -0.030)	-0.036 -(0.047 - -0.026)
PD=3	-0.052 -(0.085 - -0.019)	0.376 (0.249 - 0.499)	0.065 (0.056 - 0.074)	-0.063 -(0.081 - -0.045)	-0.074 -(0.084 - -0.063)
PD=4	-0.256 -(0.290 - -0.222)	1.193 (1.060 - 1.323)	0.155 (0.146 - 0.165)	-0.214 -(0.233 - -0.195)	-0.220 -(0.233 - -0.207)
PD=5	-0.279 -(0.315 - -0.242)	1.582 (1.448 - 1.725)	0.196 (0.185 - 0.205)	-0.265 -(0.284 - -0.245)	-0.274 -(0.288 - -0.261)
AD=2	-0.059 -(0.092 - -0.024)	0.331 (0.208 - 0.450)	0.034 (0.026 - 0.042)	-0.057 -(0.075 - -0.040)	-0.053 -(0.064 - -0.043)
AD=3	-0.102 -(0.137 - -0.066)	0.378 (0.255 - 0.507)	0.066 (0.057 - 0.075)	-0.076 -(0.094 - -0.059)	-0.089 -(0.100 - -0.078)
AD=4	-0.246 -(0.283 - -0.211)	1.352 (1.223 - 1.491)	0.168 (0.159 - 0.178)	-0.235 -(0.255 - -0.216)	-0.234 -(0.248 - -0.221)
AD=5	-0.227 -(0.260 - -0.191)	1.466 (1.329 - 1.601)	0.190 (0.180 - 0.200)	-0.240 -(0.259 - -0.223)	-0.258 -(0.271 - -0.245)
duation paramater			0.000 (0.000 - 0.000)		-0.115 -(0.120 - -0.110)
constant DCE				-5.972 -(8.717 - -3.306)	-5.756 -(8.447 - -2.991)
slope DCE				0.000 (0.000 - 0.000)	-0.080 -(0.085 - -0.075)
constant DCE_{TTO}					-0.699 -(0.946 - -0.451)
slope DCE_{TTO}					-0.407 -(0.387 - -0.428)
Deviance	11100 (11090 - 11120)	7502 (7492 - 7516)	31651 (31640 - 31670)	18612 (18600 - 18630)	50323 (50310 - 50340)
DIC	11122	7522	31671	18636	50349

Table 3. Example of EQ-5D-5L values calculated from the 20 parameter models

	TTO and DCE value set	Example: the value for health state 21111	Example: the value for health state 55555	TTO and DCE and DCE PRET-AS value set	Example: the value for health state 21111	Example: the value for health state 55555
Constant	0.893886	0.893886	0.893886	0.876663	0.876663	0.876663
MO=2	-0.04965	-0.04965		-0.02196	-0.02196	
MO=3	-0.06553			-0.04417		
MO=4	-0.17935			-0.15262		
MO=5	-0.22891		-0.22891	-0.18767		-0.18767
SC=2	-0.04343			-0.02287		
SC=3	-0.06522			-0.04428		
SC=4	-0.15044			-0.14675		
SC=5	-0.1746		-0.1746	-0.18289		-0.18289
UA=2	-0.03792			-0.03794		
UA=3	-0.04216			-0.05951		
UA=4	-0.13079			-0.12769		
UA=5	-0.13858		-0.13858	-0.1351		-0.1351
PD=2	-0.04683			-0.0365		
PD=3	-0.06276			-0.07386		
PD=4	-0.21393			-0.22015		
PD=5	-0.26462		-0.26462	-0.27419		-0.27419
AD=2	-0.05706			-0.05342		
AD=3	-0.07632			-0.08908		
AD=4	-0.23523			-0.23394		
AD=5	-0.24017		-0.24017	-0.25805		-0.25805
Health state value		**= 0.844236**	**= -0.15299**		**= 0.854703**	**= -0.161237**

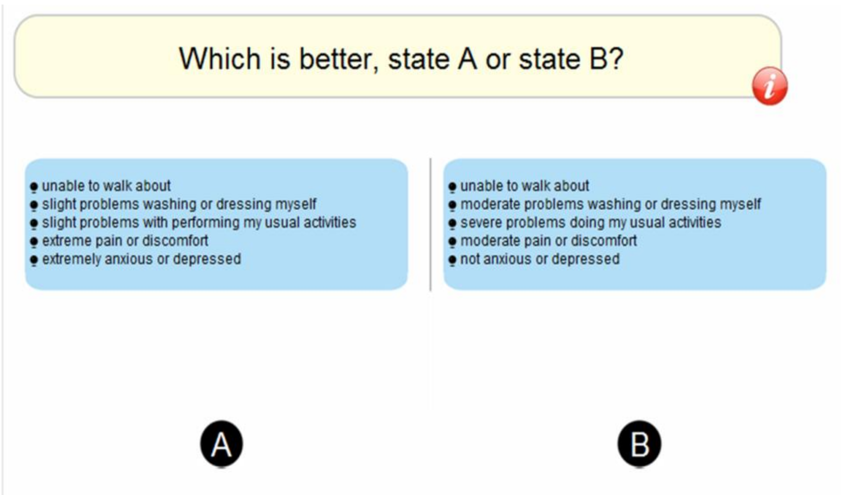
Figure 1. The composite Time Trade Off approach to valuation of health states

a. Values > 0 (better than dead) b. Values < 0 (worse than dead)



Note: As presented, Figure 1a shows a value = 0.1, and Figure 1b a value of -0.5. This screen shot from the EQ-VT is reproduced with the permission of the EuroQol Group.

Figure 2. Discrete choice task for valuation of EQ-5D-5L



Note: This screen shot from the EQ-VT is reproduced with the permission of the EuroQol Group.

Figure 3. Distribution of TTO values in the EQ-5D-5L value set for England, n = 996

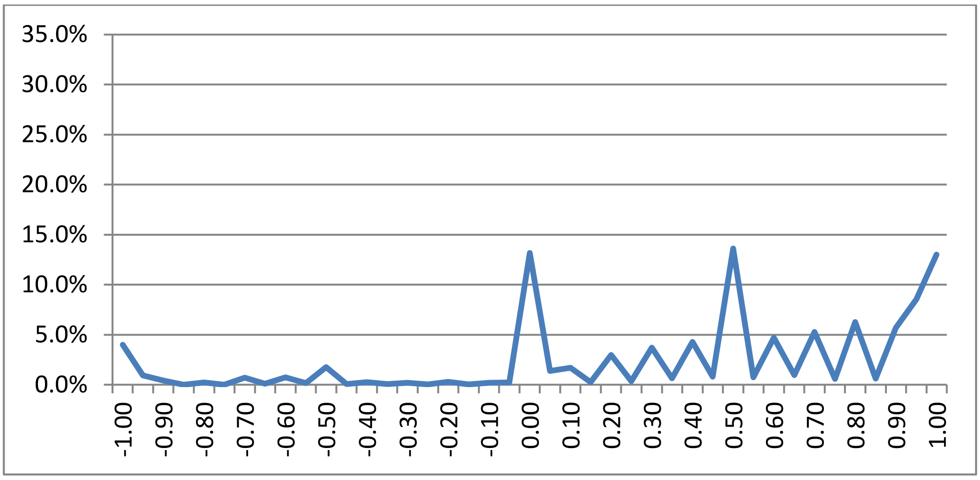
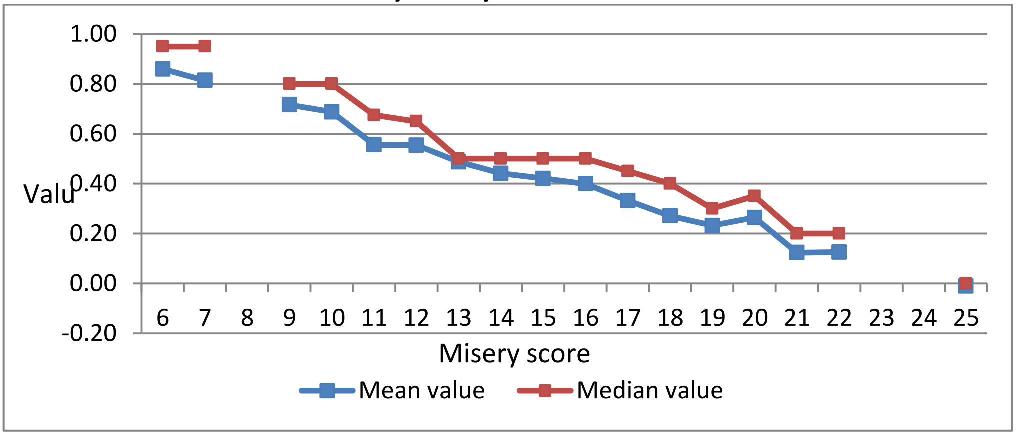
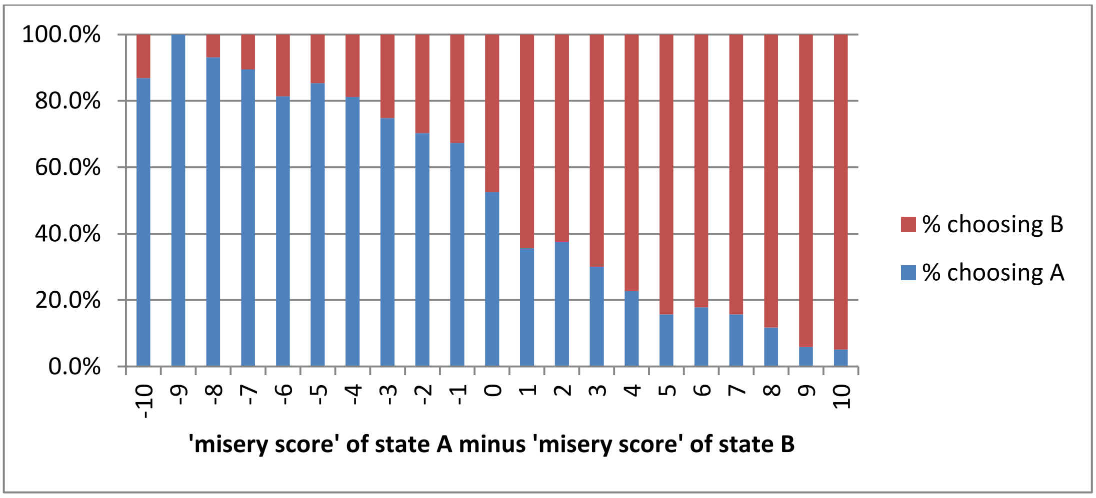


Figure 4. Mean and median TTO value by misery score



Note: The discontinuity at misery scores 8, 23 and 24 is an artefact of the optimal design underpinning the EuroQol Group research protocol.

Figure 5. Percentage choosing A and B in the DCE task, based on the relative severity of A and B



Appendix 1: Results from 5, 9 and 10 parameter models.

5 parameters	tto	dce	dce with duation	tto + dce	all
	linear model	dce logistic model	logistic model	hybrid model	hybrid model
constant	1.183 (1.152 - 1.213)			1.181 (1.152 - 1.211)	1.183 (1.149 - 1.214)
mobility	-0.048 -(0.056 - -0.040)	0.337 (0.306 - 0.369)	0.030 (0.028 - 0.032)	-0.054 -(0.059 - -0.050)	-0.058 -(0.060 - -0.055)
self care	-0.035 -(0.043 - -0.027)	0.241 (0.214 - 0.268)	0.031 (0.029 - 0.033)	-0.038 -(0.044 - -0.032)	-0.058 -(0.061 - -0.056)
uusal activities	-0.041 -(0.049 - -0.034)	0.203 (0.176 - 0.232)	0.025 (0.023 - 0.027)	-0.035 -(0.040 - -0.029)	-0.064 -(0.067 - -0.060)
pain/discomfort	-0.075 -(0.082 - -0.067)	0.406 (0.377 - 0.436)	0.046 (0.044 - 0.048)	-0.070 -(0.076 - -0.065)	-0.046 -(0.048 - -0.042)
anaxiety/depression	-0.066 -(0.074 - -0.058)	0.393 (0.364 - 0.423)	0.051 (0.048 - 0.053)	-0.068 -(0.073 - -0.062)	-0.042 -(0.045 - -0.038)
duation paramater			-0.605 -(0.626 - -0.583)		-0.607 -(0.630 - -0.584)
constant DCE				0.020 -(0.086 - 0.099)	0.526 (0.458 - 0.592)
slope DCE				5.583 (4.175 - 7.551)	-3.984 -(5.132 - -2.791)
constant DCETTO					0.102 (0.086 - 0.113)
slope DCETTO					-1.229 -(1.440 - -0.945)
Deviance	13485 (13480 - 13490)	7750 (7745 - 7757)	32154 (32150 - 32160)	21241 (21230 - 21250)	53589 (53570 - 53620)
DIC	13492	7755	32161	21249	53600

9 parameters	tto	dce	dce with duation	tto + dce	all
	linear model	dce logistic model	logistic model	hybrid model	hybri model
constant	0.876 (0.843 - 0.905)			0.880 (0.857 - 0.903)	0.858 (0.839 - 0.877)
slight	-0.181 -(0.282 - -0.103)	0.193 (0.129 - 0.342)	0.112 (0.085 - 0.150)	-0.175 -(0.296 - -0.110)	0.066 (0.048 - 0.086)
moderate	-0.340 -(0.483 - -0.235)	0.231 (0.155 - 0.412)	0.261 (0.221 - 0.357)	-0.236 -(0.395 - -0.151)	0.122 (0.091 - 0.155)
severe	-0.858 -(1.179 - -0.616)	0.710 (0.490 - 1.259)	0.728 (0.624 - 1.028)	-0.688 -(1.150 - -0.444)	0.346 (0.264 - 0.435)
unable/extreme	-0.946 -(1.295 - -0.683)	0.822 (0.571 - 1.455)	0.877 (0.755 - 1.235)	-0.776 -(1.308 - -0.504)	0.407 (0.309 - 0.512)
mobility	0.202 (0.138 - 0.284)	1.719 (0.926 - 2.391)	0.141 (0.098 - 0.164)	0.277 (0.152 - 0.397)	-0.435 -(0.568 - -0.335)
self care	0.141 (0.094 - 0.199)	1.336 (0.709 - 1.853)	0.149 (0.107 - 0.173)	0.213 (0.117 - 0.306)	-0.413 -(0.535 - -0.319)
uusal activities	0.147 (0.097 - 0.209)	1.086 (0.585 - 1.503)	0.097 (0.070 - 0.115)	0.183 (0.100 - 0.263)	-0.305 -(0.401 - -0.233)
pain/discomfort	0.284 (0.199 - 0.385)	1.960 (1.047 - 2.705)	0.212 (0.149 - 0.245)	0.334 (0.183 - 0.477)	-0.621 -(0.805 - -0.481)
anaxiety/depression	0.243 (0.169 - 0.334)	1.942 (1.033 - 2.682)	0.212 (0.149 - 0.244)	0.322 (0.177 - 0.463)	-0.599 -(0.776 - -0.463)
duation paramater			-0.386 -(0.404 - -0.369)		-0.428 -(0.446 - -0.412)
constant DCE				-0.118 -(0.173 - -0.064)	-0.116 -(0.170 - -0.062)
slope DCE				-6.499 -(6.943 - -6.075)	-6.379 -(6.769 - -6.009)
constant DCETTO					-0.081 -(0.109 - -0.053)
slope DCETTO					0.758 (0.727 - 0.794)
Deviance	11119 (11110 - 11130)	7516 (7510 - 7524)	31713 (31710 - 31720)	18628 (18620 - 18640)	50360 (50350 - 50370)
DIC	11119	7520	31721	18540	50360

10 parameters	tto	dce	dce with duation	tto + dce	all
	linear model	dce logistic model	logistic model	hybrid model	hybri model
constant	0.865 (0.837 - 0.893)			0.878 (0.856 - 0.901)	0.857 (0.839 - 0.878)
slight	-0.288 -(0.506 - -0.152)	1.134 (0.750 - 1.772)	0.192 (0.105 - 0.323)	-0.718 -(1.011 - -0.462)	-0.101 -(0.127 - -0.074)
moderate	-0.605 -(1.017 - -0.413)	1.361 (0.919 - 2.110)	0.448 (0.246 - 0.734)	-0.967 -(1.327 - -0.634)	-0.187 -(0.228 - -0.142)
severe	-1.567 -(2.528 - -1.132)	4.181 (2.905 - 6.360)	1.252 (0.675 - 2.014)	-2.827 -(3.816 - -1.891)	-0.528 -(0.646 - -0.406)
unable	-2.161 -(3.511 - -1.562)	4.772 (3.302 - 7.161)	1.516 (0.826 - 2.457)	-3.327 -(4.541 - -2.222)	-0.627 -(0.765 - -0.482)
extreme	-1.519 -(2.444 - -1.089)	4.892 (3.396 - 7.384)	1.503 (0.812 - 2.408)	-3.108 -(4.193 - -2.075)	-0.619 -(0.757 - -0.475)
mobility	0.096 (0.055 - 0.130)	0.290 (0.185 - 0.407)	0.088 (0.050 - 0.148)	0.063 (0.045 - 0.091)	0.279 (0.227 - 0.358)
self care	0.068 (0.039 - 0.094)	0.227 (0.145 - 0.321)	0.093 (0.053 - 0.157)	0.048 (0.033 - 0.069)	0.265 (0.214 - 0.340)
uusal activities	0.069 (0.039 - 0.096)	0.185 (0.117 - 0.260)	0.060 (0.034 - 0.100)	0.042 (0.029 - 0.060)	0.196 (0.156 - 0.254)
pain/discomfort	0.166 (0.097 - 0.224)	0.326 (0.208 - 0.453)	0.133 (0.075 - 0.226)	0.080 (0.057 - 0.114)	0.403 (0.325 - 0.519)
anaxiety/depression	0.145 (0.085 - 0.196)	0.324 (0.206 - 0.447)	0.133 (0.075 - 0.228)	0.077 (0.054 - 0.110)	0.389 (0.314 - 0.500)
duation paramater			-0.386 -(0.403 - -0.368)		-0.404 -(0.420 - -0.387)
constant DCE					-0.116 -(0.171 - -0.062)
slope DCE					-6.387 -(6.797 - -5.989)
constant DCETTO					-0.081 -(0.108 - -0.054)
slope DCETTO					0.759 (0.723 - 0.794)
Deviance	11104 (11100 - 11120)	7516 (7510 - 7527)	31714 (31710 - 31720)	18626 (18620 - 18640)	50361 (50350 - 50370)
DIC	11094	7523	31668	18610	50371

Appendix 2: Results from random coefficient models

	Random coefficient model	Random coefficient model with random slope on 3 points	Random coefficient model with lognormal random slope
Constant	0.898 (0.875 - 0.923)	0.864 (0.845 - 0.882)	0.892 (0.877 - 0.908)
MO=2	0.016 (18.130 - 23.800)	-0.045 -(0.233 - 0.131)	0.012 -(0.037 - 0.063)
MO=3	0.034 -(0.005 - 0.038)	0.156 -(0.092 - 0.378)	0.054 -(0.010 - 0.116)
MO=4	0.156 (0.009 - 0.060)	1.498 (1.185 - 1.799)	0.442 (0.359 - 0.538)
MO=5	0.208 (0.128 - 0.185)	2.258 (1.913 - 2.602)	0.622 (0.546 - 0.709)
SC=2	0.041 (0.180 - 0.237)	0.212 (0.001 - 0.416)	0.088 (0.034 - 0.145)
SC=3	0.051 (0.016 - 0.063)	0.392 (0.169 - 0.622)	0.116 (0.046 - 0.182)
SC=4	0.119 (0.024 - 0.075)	1.256 (0.965 - 1.563)	0.358 (0.273 - 0.448)
SC=5	0.199 (0.086 - 0.150)	2.094 (1.823 - 2.401)	0.591 (0.516 - 0.695)
UA=2	0.036 (0.171 - 0.227)	0.138 -(0.044 - 0.323)	0.063 (0.014 - 0.115)
UA=3	0.068 (0.014 - 0.058)	0.450 (0.195 - 0.727)	0.148 (0.070 - 0.227)
UA=4	0.134 (0.035 - 0.097)	1.392 (1.127 - 1.667)	0.406 (0.333 - 0.490)
UA=5	0.151 (0.105 - 0.162)	1.751 (1.456 - 2.047)	0.499 (0.417 - 0.601)
PD=2	0.028 (0.125 - 0.179)	0.100 -(0.059 - 0.271)	0.056 (0.013 - 0.099)
PD=3	0.051 (0.008 - 0.049)	0.595 (0.342 - 0.838)	0.161 (0.089 - 0.239)
PD=4	0.253 (0.024 - 0.077)	2.736 (2.423 - 3.082)	0.769 (0.669 - 0.893)
PD=5	0.297 (0.225 - 0.283)	3.188 (2.821 - 3.559)	0.908 (0.791 - 1.058)
AD=2	0.064 (0.263 - 0.329)	0.422 (0.238 - 0.596)	0.151 (0.105 - 0.206)
AD=3	0.100 (0.043 - 0.086)	0.756 (0.483 - 1.034)	0.235 (0.160 - 0.309)
AD=4	0.262 (0.069 - 0.132)	2.539 (2.216 - 2.841)	0.721 (0.633 - 0.827)
AD=5	0.240 (0.232 - 0.292)	2.519 (2.225 - 2.843)	0.702 (0.624 - 0.794)
Slope 1		0.281 (0.254 - 0.311)	
Slope 2		0.139 (0.128 - 0.150)	
Slope 3		0.048 (0.044 - 0.052)	
P(group 1)		0.116 (0.096 - 0.138)	
P (group 2)		0.366 (0.334 - 0.399)	
P (group 3)		0.518 (0.484 - 0.553)	
sigma			-1.145 -(1.251 - -1.055)
sigma scale			2.290 (2.110 - 2.502)
Pits	-0.197 -(0.215 - -0.179)	-0.358	-0.387
DIC	4746	41556	-3038