EuroQol Working Paper Series

Number 15002
October 2015

ORIGINAL RESEARCH

Valuing EQ-5D-5L using TTO and DCE: Does dimension order impact on health state values?

Brendan Mulhern[1]
Koonal K. Shah [2]
MF Bas Janssen[3]
Louise Longworth [4]

[1] University of Technology Sydney, Sydney, Australia
[2] Office of Health Economics, London, UK
[3] EuroQol Office, Rotterdam, the Netherlands
[4] Brunel University London, London, UK

EuroQol Research Foundation, Rotterdam, the Netherlands

Abstract

Objectives: Health states defined by multi attribute instruments such as EQ-5D-5L can be valued using Time Trade Off (TTO) or Discrete Choice Experiment (DCE) methods. A key feature of the tasks is the order in which the health state dimensions are presented. Respondents may use a variety of heuristics to complete the tasks, and therefore the order of the dimensions may impact on the importance assigned to particular states. The objective of this study was to assess the impact of different EQ-5D-5L dimension orders on health state values. **Methods**: Preferences for EQ-5D-5L health states were elicited from a broadly representative sample of members of the UK general public. Respondents valued EQ-5D-5L health states using TTO and DCE across one of three dimension orderings via a face-to-face computer assisted personal interview. Differences in mean values and the size of the health dimension coefficients across the arms were compared using difference testing and regression analyses. **Results**: Descriptive analysis suggested some differences between the mean TTO health state values across the different dimension orderings, but these were not systematic. Regression analysis suggests that the magnitude of the dimension coefficients differs across the different dimension orders (for both TTO and DCE), but there is no clear pattern. **Conclusions**: There is some evidence that the order in which the dimensions are presented impacts on the coefficients, which may impact on the health state values provided. The order of dimensions is a key consideration in the design of health state valuation studies.

Conflicts of interest: The authors have indicated they have no conflicts of interest with regard to the content of this article

Keywords

EQ-5D, Health state valuation, Health related quality of life, Time Trade Off, Discrete Choice

Acknowledgements

We are grateful for the contributions of the interview team, and also John Brazier, Nancy Devlin, Liz Flower, Juan Manuel Ramos-Goñi, Kim Rand-Hendriksen, Knut Stavem, Elly Stolk and Ben van Hout.

Brendan Mulhern

University of Technology Sydney
Centre for Health Economics Research and Evaluation
Building 5, Block D, Level 2
1-59 Quay St, Haymarket
NSW 2000 Sydney, Australia
E: brendan.mulhern@chere.uts.edu.au

Introduction

Generic preference based measures such as the EQ-5D [1] or SF-6D [2,3] can be used in the economic evaluation of health interventions, and inform the estimation of Quality Adjusted Life Years (QALYs) – a single metric combining quality of life and survival that is used in cost utility analysis. Preference based measures consist of two components, the health state descriptive system (a way of describing health) and a utility scale, or value set (an associated value, for each health state based on the preferences of the general population that provides the quality weight of the QALY). The EQ-5D is the most widely used generic preference based measure [4] assessing health on five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) across three response levels generating 243 health states in total. Value sets for use in cost-utility analysis (anchored on a full health (1) and dead (0) scale) have been developed internationally [5]. A new version of EQ-5D with five response levels (EQ-5D-5L; none, slight, moderate, severe, extreme/unable) has been developed [6] and studies to produce value sets for the 3,125 health states described are ongoing.

Health states can be valued using a range of preference elicitation methods, including Time Trade Off (TTO) and Discrete Choice Experiments (DCE). TTO has been used for a range of country specific EQ-5D value sets [5], including the widely used UK value set [7]. TTO requires respondents to choose between a fixed number of years in a suboptimal health state usually described by a health state classification system and a variable number of years in a comparator health state (such as 'full health'). The number of years in the comparator state is varied until indifference is reached. The recently developed protocol for the valuation of the EQ-5D-5L uses TTO alongside DCE [8,9]. DCE requires respondents to choose between multi-attribute health states described by a measure [10]. This elicits ordinal preferences that are modelled to produce utility decrements for all levels of the health dimensions, and therefore a value for each health state described, and can be anchored on the full health to dead scale if an attribute for duration of the state is included in the design [11-13]. Furthermore, TTO and DCE data can be seen as complementary preference information, and can be combined to estimate utility values. This can be done using hybrid models that combines the two sources of data using likelihood estimation [14], or by using TTO values for the best and worst health states to anchor the DCE values [15].

For both TTO and DCE there is a wide range of design specifications that can be used when framing the choice tasks which may impact on the health state valuation process [13]. One of these relates to the order in which the health dimensions are presented to respondents, and the extent to which this impacts the magnitude of the coefficients produced for each dimension is unclear. The literature suggests that the way in which people read (for example left to right; top to bottom) may lead to bias [16], and this is supported to some extent by recent eye tracking studies [17]. However, Tsuchiya and colleagues [18] presented EQ-5D-5L health states in different dimension orders in a DCE task and found no systematic impact on coefficients. There is also evidence that the relative importance of the first two EQ-5D dimensions differs depending on the type of health state valuation task used [19], and this may have implications for the choice of dimensions presented first.

The unclear findings regarding heath state dimension order suggest that it is important to systematically test the impact of reordering the dimensions on both the health state values provided, and the magnitude of the modelled coefficients for each health state dimension. Therefore the aim of this study is to assess the impact of different dimension orderings on the valuation of EQ-5D-5L health states using the EuroQol valuation protocol that combines TTO and DCE methods.

Methods

Valuation task

The EuroQol Group's recommended protocol for the valuation of EQ-5D-5L health states was used [8]. For the TTO element of the study, the composite TTO (c-TTO) [9] approach was used. The task started with the standard approach for all states [20], and moved to the Lead-Time TTO (LT-TTO) [21,22] approach when a state was valued as worse than dead. Following c-TTO for states better than dead, respondents chose between two 'lives', Life A which was a variable amount of time (t) in a 'good' health state, and Life B, which was a fixed amount of time (10 years) in a suboptimal EQ-5D-5L health state. For states worse than dead, they chose between 10 years in full health followed by 10 years in the health state (Life B) and a variable time between 0 and 20 years in full health (Life A). If an amount of time between 0 and 10 years was selected then the task remained the same, otherwise the task returned to the standard format. The time in Life A was then varied following a fixed iterative process until the respondent was indifferent between the options. At the point of indifference the TTO value for the health state can be calculated as follows:

$U = t/10$ for conventional TTO (better than dead health states)

or $U = (t - 10)/10$ for lead-time TTO (worse than dead health states)

where *U* is the value (utility) and *t* is the number of years in Life A at the respondent's point of indifference. Using the time frames described above, the possible values can range between -1 and 1. For the DCE task, respondents were presented with two EQ-5D-5L health states and asked which one they thought was better. The health state dimensions and positions of the states (i.e. whether the state appeared on the left or right of the choice set) were randomised.

Study design and state selection

A three arm study was used to test the impact of dimension ordering as part of a study that also tested the impact of varying the description of the Life A comparator in the TTO tasks (the results of which are reported by Shah and colleagues [23]). Each of the three arms administered a different EQ-5D-5L dimension ordering. The three orderings were mobility-self-care-usual activities-pain/discomfort-anxiety/depression (MO-SC-UA-PD-AD; Arm 1 - the 'standard' ordering); AD-PD-UA-SC-MO (Arm 2 - reversed ordering); and PD-AD-MO-SC-UA (Arm 3 - 'block shift' ordering), and the same order was used for both the TTO and DCE tasks. The standard order was used as a control arm, and the reverse order was selected to try to understand the change in the magnitude of the coefficients based on a complete reverse and examine potential top to bottom bias. The block shift order was used to test the impact of changing the position of the block of three 'functional' dimensions (MO-SC-UA) whilst also providing a clearly different order, and testing the impact of moving the dimensions usually appearing in the middle of the description to the first and last positions. Approximately 150 respondents were allocated to each arm which allows for an expected difference between TTO mean values of 0.15 to be detected assuming a power of 0.8, significance of 0.05, and an SD of 0.5 (in line with the mean values found in the English EQ-5D-5L valuation study).

A range of mild, moderate and severe states were used in this study (see Table 1). EQ-5D health states can be described as having sum score (a proxy for severity calculated by summing the five dimension levels; states with higher sum scores can typically be described as being more severe than those with lower sum scores), and the score was used to support the state selection, and in a number of the analyses reported here.

Three very mild states (defined as those with slight problems on one dimension and no problems on the others generating a sum score of six) were included to test the impact of varying the 'Life A' comparator (see [23] for more information). Four mild to moderate states (with a sum score between 9 and 15) were chosen from commonly occurring health states in self-reported EQ-5D-5L data taken from another preference survey carried out with the UK general population [12]. A further two severe states taken from the EQ-5D-5L EuroQol valuation protocol (sum score 18) were included. Finally, the worst state described by EQ-5D-5L was included.

The seven pairs of DCE states were based on a block of states from the EQ-5D-5L valuation study design, with two pairs (31113 vs. 11331 and 44222 vs 22244) hand selected to test the extent to which choices are driven by the dimensions appearing first and last, and first and second.

Sample selection

A target of 450 interviews was sought with South Yorkshire in the United Kingdom used as the target area. To achieve this, ten sampling points were identified, with 45 interviews targeted for each area. There are 185 towns listed in South Yorkshire, and to enable a split of areas every 18th or 19th area was selected from an alphabetical list. A central point of each town name and postcode was selected to identify a street name. The postcode and first house number was then entered into the Names and Numbers software [24] to generate a database of residential properties around this central point. The diameter of this point was expanded or contracted in order to generate the required number of addresses to enable the sample to be achieved.

Recruitment and interview process

Interviews were carried out by a team of three interviewers working for Sheffield Hallam University who were experienced in the collection of TTO and DCE data using EQ-5D. To recruit respondents, interviewers sent out letters (including an information sheet) to the selected postcode areas informing potential respondents that they would be interviewing in the area, and to contact them to organise an interview if they were interested in taking part. Interviewers then visited the sampled area to recruit respondents. The study was explained, and interested respondents signed an informed consent form. Interviews were then administered using computer assisted personal interview (CAPI) methods. Firstly, respondents completed background questions and the EQ-5D-5L for their own health. They then completed a TTO example (the wheelchair task) and three practice tasks (a mild, severe and moderate state, in that order). Following this, ten or eleven states (depending on the arm) were valued using TTO and feedback questions about the process were completed. Respondents then completed seven DCE tasks and related feedback questions. Finally, they completed a separate pen-and-paper follow-up questionnaire (not reported here). Ethical approval for the study was gained from the School of Health and Related Research, University of Sheffield ethics committee.

Data analysis

For TTO, descriptive and regression analysis was carried out on the actual TTO values provided. Post hoc, further exploratory analysis of the TTO results with all negative values censored at 0 was carried out. This was done due to the large standard deviations found in the data, and the possibility that one of the orders may have more negative values linked to a cluster effect in regard to response behaviour, where certain respondents may be more likely to provide negative values. Analysis censoring valuation data in this way has been carried out during the modelling of the English EQ-5D-5L value set development study [25].

Descriptive statistics across the three arms were compared, overall and for each health state. One-way ANOVA (with Bonferroni post hoc testing) was used to test for differences across the arms. Two-way ANOVA was used to test to assess interviewer effects across the three arms. This is important as interviewer effects have been found to impact on the characteristics of TTO data (Mulhern et al, 2013). Regression analysis was used to examine the impact of each state valued, ordering and background variables on TTO utility score. The standard model specification was:

$$y_{ij} = \alpha + \boldsymbol{\beta}\mathbf{h}_{\mathbf{j}} + \gamma o_{\mathbf{i}} + \boldsymbol{\delta}\mathbf{s}_{\mathbf{i}} + \varepsilon_{ij} \quad (1)$$

where $i = 1, 2, ..., n$ represents individual respondents and $j = 1, 2, ..., m$ represents EQ-5D-5L health states, y represents the TTO utility value, **h** represents the dummy variables for the health states or individual EQ-5D-5L health states, o represents the dummy variable for dimension order, s represents the background characteristics, and ε_{ij} represents the error term. Random effects generalised least squares regression (which takes into account multiple observations per respondent) was used to examine the magnitude of the coefficients and the relative importance of the variables across the orderings.

Given the limited number of health states included in the design, five parameter models (with one parameter for each of the dimensions) were tested including an order dummy variable and variables for background characteristics. A five parameter model with one parameter for each of the five dimensions was tested. This specification was used due to the limited number of health states included in the design. Model 1 was estimated for the overall sample, with Models 2, 3 and 4 estimated for each ordering separately. A further model (Model 5) tested the impact of each state valued, dimension order and background characteristics.

For DCE, chi square tests were used to assess differences in proportions of respondents choosing each option across the orderings. To assess the impact of ordering on the relative magnitude of the coefficients, conditional logit regression was used, with clusters specified within respondents to take into account multiple observations per respondent. Five parameter models (including one parameter for each of the five dimensions) were tested for each of the three orderings. To test for differences in the coefficient estimates from the three orders, the null hypothesis that preferences are heterogeneous across the orders was tested (following Swait and Louviere [26]). The likelihood-ratio test statistic is given by LR = -2(LL_R – LL_U) where LL_R is the log-likelihood of a model estimated on a combined sample which allows for scale differences but assumes that the value of living in full health for a specified duration (β) and the disutility of an EQ-5D-5L health state for a certain duration (λ) do not vary across the samples (the restricted model). LL_U is the sum of the log likelihoods of three conditional logit models, estimated individually (unrestricted model with variation in preferences across the designs). The models were estimated using the *clogithet* Stata module [27,28].

As the aim of the study was to test dimension ordering across the two valuation methods, models combining the TTO and DCE data were not estimated. The study design included separate blocks of states for TTO and DCE, and did not allow for data pooling.

Results

Sample and exclusion criteria

Interviews were carried out between May and October 2014. In total, 456 (19.3%) of the 2,363 households sampled resulted in fully complete interviews, with the number of interviews varying across the three interviewers used (Interviewer 1 completed 116; interviewer 2 completed 171 and interviewer 3 completed 169). Respondents who did not complete the interview in full were excluded from the analysis. Table 2 presents the allocation of the 456 respondents across the study arms and the background characteristics of the sample. A large proportion of the sample (35.9%) were aged 60 years and over and the majority were female. This is higher than in the UK general population. The standard order group (Order 1) has fewer male respondents than the reverse (Order 2) and block shift (Order 3) orderings. The self-reported health of the groups measured by the EQ-VAS score does not differ, and approximately half of the sample is in health state 11111. The time spent completing the tasks and the number of moves to complete the cTTO does not differ across the arms. We also excluded the data for 13 respondents who gave the same value in all of the TTO tasks. The DCE data was assessed for unusual response patterns (e.g. AAAAAA or ABABABA) but no evidence of this was found. Therefore all respondents were included in the DCE analysis. This is consistent with the exclusion criteria used in previous valuation studies [7,23].

TTO

Descriptive analysis

Table 3 displays the descriptive TTO results and significance levels across the three orderings for the uncensored data, and the exploratory censored analysis. At the overall level (i.e. including the three arms in the ANOVA) for the uncensored data, the difference in TTO values between the EQ-5D-5L dimension orderings is significant ($F(2,4645) = 5.01$, $p < 0.01$), with the difference occurring between the standard order and the reverse ($p=0.02$) and block shift ($p=0.02$) orders. The standard order results in a lower value for all of the moderate and severe health states, and the difference is significant for 11223 ($F(2,439) = 4.17$, $p = 0.016$) and 21232 ($F(2,439) = 4.25$, $p = 0.015$). The standard order also produces the lowest values for two of the three mild states with an aggregate score of six (21111 and 11112) although these are not significant. Regarding the censored data, there are no significant differences across the arms at the overall or health state level, indicating that the differences between the study arms are related to response behaviour of states considered worse than death.

There is evidence of interviewer effects at the overall level ($F(2,4809 = 44.31$, $p < 0.001$) but no evidence of an interaction between interviewer and order ($F(4,4811) = 1.52$, $p = 0.19$) meaning that any differences across the arms are not confounded by which interviewer collected the data. This is also the case for the exploratory analysis of the censored data where there was an overall interviewer effect (F(2,4809 = 11.51, p < 0.001), but no evidence of an interaction between interviewer and order (F(4,4811) = 1.75, p = 0.19).

A large majority of the sample agreed that the TTO was easy to understand (91.4%) and it was easy to tell the difference between the states (90.3%), and a smaller majority (63.1%) agreed that it was difficult to decide on their answers. These proportions did not significantly differ across the order groups.

Modelling

Linear regression analysis suggests that order is not statistically significant at the predefined 5% level; however the p value approached this threshold (p=0.06) (Table 4; models 1 and 5). There are similarities in the magnitude and relative ordering of the coefficients across the different ordering models, with self-care and pain/discomfort having the largest decrement, and mobility having the smallest decrement (models 2 - 4). The states with an aggregate score higher than 6 (i.e. moderate and severe states) all significantly impact on TTO values (model 5). The exploratory analysis of the censored data generated results with similar characteristics.

DCE

Descriptive analysis

Figure 1 shows the proportions of the sample choosing state A across the groups (with the states ordered by difference in aggregate score (aggregate score of state B – aggregate score of state A, with the difference shown in brackets). At the overall level the proportions choosing state A significantly differed across the orders for one DCE pair (53242 vs. 44151) but not for the other six. Again, most of the sample agreed that the task was easy to understand (88.4%) and it was easy to tell the difference between the states (95.2%), and fewer (57.1%) agreed that it was difficult to decide on their answers. This did not differ across the orders.

Modelling

Table 5 displays model coefficients across the three orderings to assess the relative importance of the coefficients based on the ordering estimated using conditional logit regression clustered by respondent to allow for multiple observations per individual. The Anxiety/depression dimension consistently has the largest decrement. There appear to be differences, but these are difficult to interpret due to the limited number of pairs included in the design. The standard errors around the estimates are somewhat large indicating that the models are unstable. Table 5 also reports the model estimated on the pooled sample, which allows for scale differences but assumes homogeneity. The likelihood ratio statistic is

11.84, and therefore the null hypothesis that the parameters are equal across the groups is rejected (demonstrating heterogeneity in preferences between the orders).

Discussion

A key methodological issue in the valuation of health states such as those from the EQ-5D [1,7] or SF-6D [2,3] is the order in which the health dimensions are presented. We have provided evidence that mean TTO values differ when the order of the EQ-5D-5L dimensions is systematically varied, some to a significant level, with the standard order producing lower values for moderate and severe states. However, exploratory analysis censoring all negative values at 0 found that the differences across the orders reduced. Regression analysis found some evidence that the magnitude of the dimension coefficients varies across the different orderings. For DCE, the magnitude of preferences seem to differ across the three order groups, but there is no systematic pattern.

The results of this study suggest that dimension ordering may have some impact on health state valuation, and raise questions about whether future valuation studies should impose dimension randomisation between subjects. This was employed in a recent online valuation of EQ-5D using TTO [29] and is being applied in ongoing international valuation studies of SF-6D-V2 using DCE incorporating survival [30]. If order effects are important, then randomisation has the advantage of averaging the effect out. Furthermore, no one order seems to be more difficult than the others (based on respondent feedback) so randomisation within the health state valuation process is feasible and could be considered in the design of future studies alongside the range of methodological choices that are made. It is worth noting that there would be possible issues with the comparability of values across previously completed and new valuation studies using the same instrument if randomisation was introduced.

However, the results of this study imply that the impact of changing the dimension order is not systematic. This may be related to state severity and associated data characteristics including the wider range of values and a larger amount of worse than dead valuations for more severe states. More evidence about the nature of the impact and how it differs across different valuation methods with different experimental design issues is required. Regarding EQ-5D, the 'standard' order was originally used as it was hypothesised to allow respondents to build up a 'narrative' picture of any health state starting with the functional dimensions (mobility, self-care, usual activities), and using a ordering that makes the valuation process as amenable to respondents as possible may be a consideration.

It is also worth considering the potential impact of dimension randomisation could have on TTO values produced, and the subsequent modelled values. A key finding of this study is the lower TTO values for non-mild states using the standard EQ-5D dimension order. This suggests that differences in the value sets already produced would be possible if randomisation had been used, particularly as in some cases the difference in mean values is larger than the minimally important difference suggested for EQ-5D to detect significant change [31]. However, the difference between study arms is much smaller when censoring the negative values at 0, which indicates that respondents in the standard order arm on average valued more states as worse than death and traded more life years for these states. Valuation data revealed different types of response behaviour particularly regarding the valuation of states worse than dead [32], and it might be a chance finding that explains the lower values in the standard order arm.

The results of this study must be considered in the context of a range of limitations. Firstly due to constrictions related to the design we were only able to test three of the possible 120 EQ-5D-5L orders. However we believe that the orders were guided by clear hypotheses tested provide valid comparisons to each other in terms of the position of the dimensions. We were also only able to value a small number of states, and this has implications for the interpretation of the coefficients for both the TTO and DCE models, which would be more informative with more health states included, and may inform the pattern of coefficient differences. We also did not use a formal experimental design. Further research could test a wider range of orders with a larger sample size and number of states (selected using a formal experimental design). This will help to establish the extent to which dimension order is a key consideration in comparison to other experimental design issues such as the state selection procedure used. The impact of dimension other on other methods such as DCE incorporating duration should also be investigated. There is a particular concern about the modelled coefficients for mobility, as these differ from the other dimensions. However, the states were chosen to cover the full severity range, and the combined models for TTO and DCE have sufficient power and provide an indication that order impacts the relative size of the dimension coefficients, and that preferences differ across the order groups.

References

1 Brooks R. EuroQol: The current state of play. Health Pol 1996;37:53-72.

2 Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002;21:271-92.

3 Brazier JE, Roberts J. Estimating a preference-based index from the SF-12. Med Care 2004;42(9):851-59.

4 Wisløff T, Hagen G, Hamidi V, et al. Estimating QALY gains in applied studies: a review of cost-utility analyses published in 2010. Pharmacoeconomics 2014;32:367-75.

5 Szende A, Oppe M, Devlin N. EQ-5D valuation sets: an inventory, comparative review and users' guide. Rotterdam: Springer, 2007.

6 Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011;20(10):1727-36.
7 Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35(11):1095-108.
8 Oppe M, Devlin NJ, van Hout B, et al. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value Health 2014;17:445-53.
9 Janssen BM, Oppe M, Versteegh M, Stolk E. Introducing the composite time trade-off: a test of feasibility and face validity. European Journal of Health Economics 2013;14(1):5-13.
10 Stolk EA, Oppe M, Scalone L, Krabbe P. Discrete choice modeling for the quantification of health states: the case of the EQ-5D.Value Health 2010;13(8):1005-13.
11 Bansback N, Brazier J, Tsuchiya A, et al. Using a discrete choice experiment to estimate health state utility values. J Health Econ 2012;31(1):306-18.
12 Mulhern B, Bansback N, Brazier J, et al. Preparatory study for the re-valuation of the EQ-5D tariff: Methodology report. Health Technol Assess 2014;18:12.
13 Norman R, Cronin P, Viney R. (2013) A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. Applied Health Econ Health Pol 2013;11(3):287-98.
14 Ramos-Goñi JM, Pinto-Prades JL, Cabasés JM, Rivero-Arias O. Valuation and modeling of EQ-5D-5L health states using a hybrid approach. Med Care 2014;doi: 10.1097/MLR.0000000000000283.
15 Rowen D, Brazier J, van Hout B. A comparison of methods to convert DCE values onto the full health – dead QALY scale. Medical Decision Making 2014;35(3): 328-40.
16 Spalek TM, Hammad S. The left-to-right bias in inhibition of return is due to the direction of reading. Psychol Science 2005;16(1):15-18.
17 Krucien N, Ryan M, Hermens F. Using Eye-tracking methods to inform decision making processes in Discrete Choice Experiments. Glasgow: Health Economist's Study Group, 2014.
18 Tsuchiya A, Mulhern B, Bansback N, et al. Using DCE with duration to examine the robustness of preferences across the five dimensions of the EuroQol instrument: The second paper from the FEDEV project. Stockholm: EuroQol Group Plenary, 2014.
19 Rand-Hendriksen, K, Augestad LA. Time Trade-Off and Ranking Exercises Are Sensitive to Different Dimensions of EQ-5D Health States. Value Health 2012;15(5):777–82.
20 Brazier J, Ratcliffe J, Salomon J, Tsuchiya A. Measuring and valuing health benefits for economic evaluation. Oxford: University Press, 2007.
21 Devlin N, Tsuchiya A, Buckingham K, et al. A uniform Time Trade Off method for states better and worse than dead: feasibility study of the 'lead time' approach. Health Econ 2011;20(3):348-61.
22 Devlin N, Buckingham K, Shah K, et al. A comparison of alternative variants of the lead and lag time TTO. Health Econ 2013;22(5):517-32.
23 Shah K, Mulhern B, Longworth L, Janssen MF. An empirical study of two alternative comparators for use in time-trade off studies. EuroQol Working Paper 15001. Rotterdam: EuroQol Research Foundation, 2015.

24 AFD Software. Names and numbers software 2013 (available at http://www.afd.co.uk/product_namesandnumbers.asp)
25 van Hout B, Devlin N. Developing an EQ-5D-5L value set for England. Office of Health Economics seminar, London; 2014.
26 Swait J, Louviere J. The role of the scale parameter in the estimation and comparison of multinomial logit models. Journal of Marketing Research 1993;30(3):305-14.
27 Hole AR. CLOGITHET: Stata module to estimate heteroscedastic conditional logit models. Statistical Software Components S456737, Boston College Department of Economics, 2006.
28 Hole AR. Small-sample properties of tests for heteroscedasticity in the conditional logit model. Economics Bulletin 2006;3(18):1-14.
29 Bansback N, Tsuchiya A, Brazier JE, et al. Canadian Valuation of EQ-5D Health States: Preliminary Value Set and Considerations for Future Valuation Studies, PLoS ONE 2013;7(2):e31115.
30 Mulhern B, Brazier J. Developing SF-6Dv2: The health state classification system. Qual Life Res 2014;23:49.
31 Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res 2005;14(6):1523-32.
32 Oppe M. Cluster analysis of Time Trade Off response behaviour. 2014

Table 1: States used in study (described in standard EQ-5D ordering)

TTO states	DCE state pairs	
	Life A	Life B
11111	42525	53422
11112	53242	44151
11121	31113	11331
11223	44222	22244
21111	44241	15244
21232	22331	22413
32442	23144	42452
34155		
43331		
55233		
55555		

Table 2: Respondent characteristics across the order groups

	Overall	MO-SC-UA-PD-AD (Standard)	AD-PD-UA-SC-MO (Reverse)	PD-AD-MO-SC-UA (Block shift)	Sig
N	456	151	155	150	
Overall					
11111 comparator	231	76	77	78	
Full health comparator	224	74	78	72	
Male n(%)	195 (42.7)	53 (35.1)	76 (49.0)	66 (44.0)	0.04
Age					
M(sd)	50.1 (18.2)	49.6 (18.1)	50.6 (18.1)	50.2 (18.4)	0.90
Range	17-93	17-82	18-93	18-86	
Experience of illness					
Caring for others n(%)	174 (38.2)	62 (41.3)	57 (36.8)	55 (36.7)	0.64
In family n(%)	293 (64.4)	93 (62.0)	100 (64.5)	100 (66.6)	0.70
Yourself n(%)	155 (34.1)	54 (36.0)	51 (32.9)	50 (33.3)	0.83
VAS score M(sd)	80.5 (18.1)	78.9 (19.6)	80.8 (18.0)	81.8 (16.4)	0.36
Respondents in 11111 n(%)	232 (51.0)	79 (52.7)	76 (49.0)	77 (51.3)	0.81
Interview duration (mins, M(sd))	23.2 (9.0)	22.4 (8.2)	23.5 (8.5)	23.7 (10.1)	0.44

Key: M (mean); sd (standard deviation)

Table 3: TTO descriptive statistics (uncensored and censored values)

State[1]	MO-SC-UA-PD-AD (n=146) (Standard)		AD-PD-UA-SC-MO (n=149) (Reverse)		PD-AD-MO-SC-UA (n=148) (Block shift)		Significance	
	Uncensored M(sd)	Censored M(sd)	Uncensored M(sd)	Censored M(sd)	Uncensored M(sd)	Censored M(sd)	Overall Uncensored	Censored
Overall	0.554 (0.60)	0.662 (0.39)	0.610 (0.55)	0.690 (0.38)	0.610 (0.54)	0.690 (0.37)	0.01	0.06
11111	1	1	0.998 (0.01)	0.998 (0.01)	0.999 (0.01)	0.999 (0.01)	0.36	0.36
21111	0.949 (0.14)	0.950 (0.13)	0.961 (0.11)	0.954 (0.14)	0.968 (0.10)	0.969 (0.10)	0.38	0.40
11121	0.968 (0.09)	0.969 (0.09)	0.967 (0.10)	0.959 (0.13)	0.963 (0.10)	0.964 (0.10)	0.90	0.75
11112	0.941 (0.20)	0.950 (0.14)	0.956 (0.12)	0.949 (0.15)	0.948 (0.13)	0.949 (0.12)	0.71	0.99
11223	0.757 (0.43)	0.805 (0.28)	0.848 (0.26)	0.852 (0.24)	0.849 (0.19)	0.853 (0.19)	0.02	0.13
21232	0.754 (0.38)	0.786 (0.28)	0.838 (0.24)	0.839 (0.23)	0.837 (0.20)	0.841 (0.20)	0.01	0.07
43331	0.516 (0.55)	0.615 (0.36)	0.590 (0.52)	0.670 (0.34)	0.637 (0.35)	0.660 (0.30)	0.10	0.30
32442	0.323 (0.59)	0.467 (0.36)	0.418 (0.54)	0.524 (0.36)	0.404 (0.54)	0.515 (0.35)	0.28	0.34
55233	0.238 (0.62)	0.424 (0.36)	0.310 (0.61)	0.471 (0.38)	0.287 (0.60)	0.449 (0.35)	0.59	0.54
34155	0.129 (0.62)	0.350 (0.34)	0.163 (0.63)	0.377 (0.37)	0.160 (0.61)	0.363 (0.35)	0.87	0.79
55555	-0.260 (0.58)	0.136 (0.27)	-0.144 (0.54)	0.155 (0.29)	-0.155 (0.58)	0.170 (0.28)	0.15	0.58

1 Each state is described using standard EQ-5D-5L ordering (MO-SC-UA-PD-AD)

Table 4: TTO regression - Uncensored data

	Model 1: All (level)		**Model 2: Standard**			**Model 3: Reverse**			**Model 4: Block shift**			**Model 5: All (state)**	
	Coef.	Sig.	Coef.	Sig	Order	Coef.	Sig		Coef.	Sig		Coef.	Sig
MO	-0.002	0.87	-0.029	0.32	5	0.012	0.69	5	0.008	0.75	5		
SC	-0.105	0.00	-0.078	0.01	1	-0.120	0.00	1	-0.116	0.00	1		
UA	-0.061	0.00	-0.070	0.00	4	-0.057	0.00	3	-0.057	0.00	4		
PD	-0.073	0.00	-0.074	0.00	2	-0.074	0.00	2	-0.071	0.00	2		
AD	-0.059	0.00	-0.071	0.00	3	-0.048	0.00	4	-0.058	0.00	3		
State													
11112												-0.036	0.20
11121												-0.019	0.51
11223												-0.166	0.00
21111												-0.025	0.37
21232												-0.175	0.00
32442												-0.603	0.00
34155												-0.833	0.00
43331												-0.403	0.00
55233												-0.706	0.00
55555												-1.171	0.00
Order	0.029	0.06										0.029	0.06
Gender	0.034	0.19	0.086	0.11		-0.016	0.68		0.048	0.21		0.034	0.19
Age	-0.001	0.09	-0.001	0.31		-0.002	0.14		-0.000	0.69		-0.001	0.09
N	443		146			149			147			443	
Wald	5615		1988			1722			1927			5765	

Table 5: DCE regression

	MO-SC-UA-PD-AD (Standard)				AD-PD-UA-SC-MO (Reverse)				PD-AD-MO-SC-UA (Block shift)				Heteroscedastic model		
	Coef.	Sig.	SE	Rank	Coef.	Sig.	SE	Rank	Coef.	Sig.	SE	Rank	Coeff	Sig	SE
MO	-0.198	0.05	0.08	4	-0.264	0.00	0.09	2	-0.282	0.00	0.09	4	-0.227	0.00	0.065
SC	-0.281	0.04	0.11	3	-0.260	0.02	0.11	3	-0.410	0.00	0.13	2	-0.288	0.00	0.084
UA	-0.314	0.00	0.04	2	-0.241	0.00	0.04	4	-0.227	0.00	0.04	5	-0.236	0.00	0.050
PD	-0.142	0.28	0.10	5	-0.205	0.06	0.11	5	-0.394	0.00	0.12	3	-0.226	0.00	0.076
AD	-0.367	0.00	0.07	1	-0.404	0.00	0.07	1	-0.525	0.00	0.08	1	-0.394	0.00	0.077
Scale													0.044	0.59	0.081
N obs	2100				2170				2100				6370		
LL	-683				-710				-674				-2074		

Figure 1: DCE proportions based on aggregate score difference between the health states