Using a discrete choice experiment to estimate societal health state utility values

Bansback, Nick and Brazier, John and Tsuchiya, Aki and Anis, Aslam

The University of Sheffield

2010

HEDS Discussion Paper 10/03

ScHARR
School of Health and
Related Research

USING A DISCRETE CHOICE EXPERIMENT TO ESTIMATE SOCIETAL HEALTH STATE UTILITY VALUES

Nick Bansback[1,2], John Brazier[1], Aki Tsuchiya[1,3], Aslam Anis[2,4]

1. School of Health and Related Research, University of Sheffield, Sheffield, S Yorks, UK
2. Centre for Health Evaluation and Outcome Sciences, St Paul's Hospital, Vancouver, BC, Canada
3. Department of Economics, University of Sheffield, S Yorks, UK
4. School of Population and Public Health, University of British Columbia, BC, Canada

Abstract

In this study we explore a novel application of the Discrete Choice Experiment (DCE) that resembles the Time Trade Off (TTO) task to estimate values on the health utility scale for the EQ-5D. The DCE is tested in a survey alongside the TTO in respondents largely representative of the Canadian general population. The study finds that the DCE is able to derive logical and consistent values for health states valued on the full health – dead scale. The DCE overcame some issues identified in the version of TTO currently used to value EQ-5D, notably whether to exclude respondents who fail to understand the task and incorporating values considered worse than dead without transformation. This has important implications for providing values that represent the preferences of all respondents.

Acknowledgements: We are grateful to Huiying Sun and Daphne Guh for their technical support, discussants and participants of the HESG in Sheffield 2009, and EuroQol group 2009 meetings for their comments on earlier versions of the paper. The study reported in this study is partly funded by the Canadian Institute for Health Research. The usual disclaimer applies.

Corresponding author: Nick Bansback, Centre for Health Evaluation and Outcomes Sciences, St Paul's Hospital, Vancouver. Tel: 1-604-682-2344 (63561) Email: nbansback@cheos.ubc.ca

Key words: Preference-based health measures; EQ-5D; health state valuation, DCE, TTO

1. Introduction

The Quality Adjusted Life Year (QALY) is a widely used measure of health improvement for guiding health-care resource allocation decisions. A key input to QALY calculations is the relative value of time spent in different health states (Torrance 1986). Methods for eliciting these values have been dominated by cardinal preference techniques such as the Standard Gamble (SG) and Time Trade Off (TTO) giving values anchored on 1 (full health) and (0) dead (herein referred to as the "health utility scale") (Brazier et al. 2007). QALYs are to date most commonly used in societal resource allocation decisions, and so values are typically obtained from a representative sample of the society's population. There is concern that the tasks involved in the SG and TTO are too complex for certain populations, resulting in many inconsistencies and subsequent exclusions that limit representativeness of the values (Craig et al. 2009).

As a consequence, researchers have sought alternative elicitation methods to value health states, with ordinal techniques such as ranking becoming the focus of attention in the recent literature (e.g. Salomon 2003, McCabe et al. 2006, Ratcliffe et al. 2009, Craig et al.2009). Such techniques require respondents simply to rank responses, such as stating that health state A is preferred to B, without going through an iterative process of identifying *the degree* by which A is preferred to B. Being cognitively simpler, the choices are less prone to error and through greater inclusion, values will be more representative of all the surveyed respondents.

An alternative ordinal elicitation method that has become popular in the health economics literature is the pair wise discrete choice experiment (DCE) (Louviere et al. 2000). Typically, this approach involves the construction of sets of profiles[a] based on a descriptive system made up of levels of a limited number of important attributes. Preferences over two or sometimes more profiles are obtained by respondents simply choosing their most or least preferred. The exercise can be repeated with different profiles in order to infer the relative weight attached to each level of each attribute. By requiring individuals to trade-off between attributes, DCEs overcome some limiting assumptions in ranking data and have been shown to be consistent with the conditional logit model (Louviere and Woodworth 1983), rooted in Random Utility Theory (Mcfadden 1974). The relative simplicity of the task involved means that, in contrast to the TTO and SG, DCEs are typically conducted without an interviewer, in the past often by

[a] While a profile might be made up of a solely a health state, for this paper we refer to profiles if it includes additional attributes such as life years

paper, but more recently using computers which through the use of the internet enables fast, flexible and precise surveys.

While the appeal of using DCEs as an alternative to conventional techniques appears to have some methodological and theoretical basis, care needs to be taken to understand the limitations that DCEs bring over conventional elicitation techniques (Bryan and Dolan 2004, Lancsar and Donaldson 2005). To date, there has been little empirical research comparing DCEs to cardinal elicitation techniques and so advantages are largely theoretical. A key challenge to the use of DCEs is the anchoring of values to the health utility scale. DCE data, through the variations of the conditional logit model, can provide estimates of cardinal utility functions from ordinal preferences on the latent utility scale. While this can provide information on the relative preference of one health state to another, the scale is not anchored on full health and dead and so cannot be directly incorporated into QALY calculations.

This paper explores a new application of the DCE which closely resembles the TTO to produce health state values on the health utility scale for the EQ-5D. The DCE is tested in an on-line survey alongside the TTO in respondents largely representative of the Canadian general population. In section 2, a brief review of previous DCE studies used to value health states is presented. Section 3 describes the survey methods while section 4 details the econometric modelling and rescaling assumptions utilized in the study. The results of the TTO and DCE are described and compared in section 5. Finally the implications of these results for future elicitation of health state values are discussed.

2. A brief review of the use of DCEs to value health status

While the use of DCEs in valuing preferences for health states dates back over 10 years (e.g. Hakim and Pathak 1999), only recently have studies endeavoured to anchor the resulting values on the health utility scale for use in QALY calculations.

The studies by Ryan et al. (2006) and Burr et al. (2007) used profiles made up solely of health state descriptions and assume that the best health state in their descriptive system (level 1 of each attribute) is equivalent to full health (e.g. equal to 1 on the health utility scale) and the worst health state is equivalent to dead (e.g. equal to 0). The values for other health states are then rescaled correspondingly. This is similar to conventional valuation studies using TTO, where, for example, EQ-5D state 11111 is assumed to be equivalent to full health. However, assuming the worst health state is equivalent to dead, as discussed by the authors of the studies, is a more undesirable assumption. Studies using the SG and TTO have shown that the health state individuals consider equal to dead varies based on the descriptive system (Brazier et al. 2007) and indeed many health states are considered worse than dead.

The study by Ratcliffe et al. (2009) addresses this issue by using values for best and worst health states from the TTO to inform the rescaling. While this scaling improves the theoretical basis for the values, the reliance on using TTO elicitations contradicts the primary motivation of using DCEs instead of conventional techniques to value health states.

Another approach is to include dead in the design, thereby eliminating the reliance on external elicitations or assumptions (Flynn et al. 2008, Brazier et al. 2009). Modelling enables coefficients for attribute levels for states 'worth living' to be estimated as their distance from 'dead'. These methods rely on at least some respondents indicating that at least some states are worse than dead (or not 'worth living'). Flynn *et al*. (2008) has pointed out, however, that if a certain proportion of respondents categorically do not accept that there are such states (or that all states are 'worth living'), then this will violate the assumption that health states can be located on a continuous scale that could be anchored at dead and full health.

An alternative approach has been proposed an attribute, such as probability of death or years of survival is included as an attribute (Ryan et al. 2006, Viney et al.2007, Coast et al. 2008, Flynn et al. 2008). In the case of incorporating years of survival, this results in a DCE that resembles the TTO (the TTO could in fact be considered as a form of DCE) by asking respondents to choose between health profiles which contain a health state description and a length of life that would be lived in that health state. Since health is typically defined as the product of health status and life years, this would indicate the need for a multiplicative design in which interactions between each health status level and life years are estimated. It is this innovative design of DCE (herein referred to as DCE_{TTO} representing its link to the TTO) that is considered in this paper.

3. Survey methods

3.1 Survey and elicitation tasks

A web survey was conducted asking respondents to complete a series of TTO and DCE_{TTO} tasks. Health states in the survey were described using the EQ-5D descriptive system (Brooks 1996), which consists of five attributes (mobility, self-care, usual activities, pain/discomfort and anxiety/depression), with three possible levels for each attribute. Level 1 refers to the best level in each attribute – so health state 11111 refers to full health – and 33333 refers to the worst health state possible in the descriptive system.

The survey began by asking respondents to describe their own health using the EQ-5D. For the second stage of the survey, respondents were randomly assigned to either the TTO or DCE_{TTO} "exercise". In each, a description of the "task" involved was given, and a demonstration video provided. In both exercises, the order of attributes was randomized for each respondent (apart from 'life years' in the DCE_{TTO} which was at the bottom). At the end of the first exercise, respondents were asked about any difficulties with understanding and answering the tasks. In the third stage of the survey, the exercise not utilised in the second stage was used, followed again by questions on their difficulty with understanding and answering the task.

The TTO was originally designed as a simpler alternative to the SG for valuing health states and is reviewed in detail elsewhere (Brazier et al. 1999). Its premise is that the welfare change associated with a decrement in health status is determined by valuing the amount of life expectancy an individual is prepared to sacrifice that leaves overall utility unchanged. In this survey, we used a web version of the TTO-prop method developed by the York Measurement and Valuation Health Group (Gudex 1994). Respondents were given a hypothetical EQ-5D health state at a time and asked to assume that the duration of the health state was 10 years, followed by immediate death. A choice is then presented: to live in the given health state or to die immediately. If living in the health state is chosen, tradeoffs are made using a visual board to determine the number of years (herein denoted as *t*) in full health that is equivalent to 10 years in the state presented. If the respondent prefers immediate death over the health state presented, the survey proceeds by asking if the respondent prefers 10-*t* years in the given health state followed by *t* years in full health, or again immediate death. For both cases, the time *t* starts at 5 and is then varied by 1 year and then 6 monthly intervals, based on responses, until the point of indifference is found. To identify potentially problematic respondents, in the preliminary task of this exercise, respondents are asked whether they would prefer living in full health for 10 years or the given health state for 10 years.

The DCE_{TTO} was designed to reflect the TTO, in terms of both the task description and the instructions used but after pilot testing, visual aids were not used. Pairwise choices were developed from profiles including a health state made up of the 5 EQ-5D attributes, and a sixth attribute describing the number of years the individual would live in that health state followed by immediate death. Four levels of life years were chosen: 10, 7, 4, and 1 years. The respondent was asked to simply choose which profile they preferred (e.g. a forced choice). In addition, to test for logical consistency, two tasks were included where one of the two profiles was regarded as a dominant option (all attribute levels were regarded as more desirable and the levels of life years were equal).

A market research company was hired to recruit a representative sample of the Canadian population over 18 years of age. Initial contact was made via email. Individuals choosing to participate in the study were referred to a password-protected website that contained the survey. The market research company offered incentives to participants who completed the survey questionnaire. Socio-demographic profiles of all participants invited to the survey were provided by the market research company. Ethical approval was obtained from the University of British Columbia Ethics Board.

3.2 Experimental design

The EQ-5D has 3^5 (243) combinations of attribute levels in the full factorial design. Although smaller orthogonal arrays exist, a near orthogonal array of 36 states was used (Kuhfeld 2009) so as to allow more comparisons with health states from the DCE_{TTO}. Furthermore, it was decided to add 12 more health states (total =49) so that all 17 health states recommended by Lamers et al. (2006) were directly valued enabling the comparison of TTO with many previous valuation studies.

Including the life years' attribute to the DCE profile increases the number of combinations of attribute levels to 972, and 471,906 potential pairwise combinations for the DCE_{TTO} task. We constructed a fractional design using a D-optimality algorithm (Kuhfeld 2005) which considered the inclusion of the two–way interactions between each of the EQ-5D attributes with the Life years attribute. This procedure was complicated by the fact that the EQ-5D includes implausible attribute-level combinations (e.g confined to bed but no problems with usual activities) and that many of the developed designs included tasks that were dominated (and hence provide little information). We therefore generated a further 1000 designs based on near orthogonal arrays considering all possible combinations of attribute levels. The final design was chosen from these by comparing which design had i) the smallest correlation between specified effects (so that the covariance between attribute effects was minimized), ii) the fewest dominated pairwise profiles (which were manually altered to become non dominated), iii) the highest overlap where both profiles include health states included in the TTO (to enable further comparison in a follow on study), iv) and the highest efficiency (defined as the determinant of the Fisher Information Matrix). In the end, a design which included 117 EQ-5D health states across 144 pairwise tasks was selected, and tested using simulated data to ensure that a model could be estimated.

A small on-line pilot study was undertaken in advance of the main study to check that respondents understood the tasks, answered the questions as expected and to get feedback on the design of the website. The pilot study also suggested that in 20 minutes, a given participant could reasonably answer 5 tasks in the TTO exercise *and* 8 tasks in the DCE_{TTO} exercise. Consequently, the TTO health states were blocked into 12 sets. In each set, respondents valued the worst health state (33333) and 4 other states selected by a computer algorithm so that near level balance was achieved between sets. The 144 DCE_{TTO} pairs were also blocked into 24 sets using an algorithm which also matched each set with one of the TTO sets (e.g. 2 DCE_{TTO} sets to each TTO set) such that there was overlap where possible between states in the two sets.

4. Modelling health state values

4.1 Time Trade Off

The modelling of the TTO replicates previous studies (e.g. Dolan 1997, Lamers et al. 2006). For states valued better than dead, values are calculated by dividing the number of years in full health t (at the point of indifference) by 10 (the maximum time). For states valued worse than dead, values are calculated using a monotonic transformation so they are bounded to -1 (Dolan 1997).

A one way error components random effect model which takes into account the variation both within and between respondents (Brazier, et al. 2007) is defined as:

$$V_{ij} = f(\beta' \boldsymbol{x}_{ij}) + \varepsilon_{ij} \tag{1}$$

where i = 1, 2,...,n represents individuals and j = 1, 2,...,m represents the different health states shown to each respondent. The dependent variable, V_{ij} is the disutility value (1- mean TTO value) for health state j valued by each respondent i. $\boldsymbol{x}_{ij}$ is a vector of 10 binary dummy explanatory variables ($x^{\delta\lambda}$) where λ=2,3 indicates the levels 2 and 3 of each attribute δ=1,2,…,5 in health state. Level λ=1 reflects no problems in each attribute. Hence $\boldsymbol{\beta}$ is vector of 10 variables (β^{12}, β^{13},…, β^{53}). Finally ε_{ij} is an error term which represents the respondent-specific variation and the error term for the jth health state valuation of the ith individual, assumed to be random across observations. A linear additive function is assumed, as is commonly done. The normality of residuals and predicted random effects are assessed via graphical means. Belsey's condition index (Belsley et al. 1980) was used to assess multicollinearity and Ramsey's regression specification error test (RESET test) was used to test for functional form (Ramsey 1969). Predicted values of V_j (say $\hat{V}_j^{TTO}$) is the average value for health state $\boldsymbol{x}_j$ for a given TTO time horizon estimated directly on the health utility scale (10 years in this case).

4.2 Discrete Choice Experiment

4.2.1 Modelling framework

To model the health state valuations using the DCE_{TTO} data, we used the conditional logistic model as outlined by McFadden (1974). This operationalises the Random Utility Model, which is described below as its theory is useful in the later discussion.

The approach assumes that in a DCE, each individual *i* has a utility function for profile *k* defined as $U_{ik} = \mu_{ik} + \varepsilon_{ik}$ where μ_{ik} is an observable component - the part of the utility contributed by attributes and ε_{ik}, is a random component as it is assumed one cannot fully observe the set of influencing factors in a person's decision process. It is assumed that in a set of tasks *A*, individual *i* will choose profile *k* if and only if $U_{ik} > U_{ij}$ for all j≠k in *A*. Since $(\varepsilon_{ik} - \varepsilon_{ij})$ cannot be directly observed, it cannot be determined if $(\mu_{ik} - \mu_{ij}) > (\varepsilon_{ik} - \varepsilon_{ij})$. Instead, only the probability that $(\varepsilon_{ik} - \varepsilon_{ij})$ will be less than $(\mu_{ik} - \mu_{ij})$ can be inferred from the choices such that the probability of individual *i* choosing profile *k* is:

$$P_{ik} = P(U_{ik} > U_{ij}) = P\left\{(\mu_{ik} - \mu_{ij}) > (\varepsilon_{ik} - \varepsilon_{ij})\right)\} \quad all\ j \neq k \tag{2}$$

The conditional logit model restricts all ε_{ij} to be independent and identically distributed and exhibits an extreme value. The probability that individual *i* chooses health profile *k* can be solved as a closed form solution of:

$$P_{ik} = \frac{exp(\mu_{ik})}{\sum_{j=1}^{J} exp(\mu_{ij})} \tag{3}$$

4.2.2 Model specification

For the EQ-5D attributes in DCE_{TTO}, a similar model specification is made to equation 1 used for TTO. However we expect that each individual *i*'s utility function μ_{ij} is multiplicative between the health state and number of life years in each profile *j*. The full model can be written as:

$$\mu_{ij} = \alpha_i + \beta_1' \boldsymbol{x}_{ij} + \beta_2 t_{ij} + \beta_3' \boldsymbol{x}_{ij} \cdot t_{ij} \tag{4}$$

Estimates of $\boldsymbol{\beta}_1$ (i.e. $\hat{\beta}_1^{12}, \hat{\beta}_1^{13}, ..., \hat{\beta}_1^{53}$) are the weight associated with the level of the attribute in each health state $\boldsymbol{x}_{\mathbf{j}}$ where the $x^{\delta\lambda}$ variables are defined similarly to equation 1. The estimate of β_2 ($\hat{\beta}_2$) is the weight associated with the 'life years' attribute *t*. For consistency with the TTO analysis, respondents are assumed to have a constant proportional time trade off and the assumptions of the QALY model, and so t is considered to be linear. Each estimate of $\boldsymbol{\beta}_3$ ($\hat{\beta}_3$) is the weight associated with the time t lived in each health state $\boldsymbol{x}_j$. In the full model, we would expect all $\hat{\beta}_1$s to be equal to zero as otherwise it would indicate individuals having a preference for health states independent of time violating a key assumption of the QALY model. In reality, the inclusion of both sets of terms ($\boldsymbol{x}$ and $\boldsymbol{x}.t$) would likely cause multicollinearity in the estimation of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_3$, and so from a theoretical and estimation point of view, $\boldsymbol{\beta}_1$ is excluded from the model.

4.2.3 Anchoring to health utility scale

In the conditional logit model, the predicted value $\hat{\mu}_k$ can be interpreted as an estimate of the utility U_k of profile k. This estimate is however on the latent scale and so a further assumption is required to anchor the estimates to the health utility scale. For this, we simply borrow a method from TTO by determining the life expectancy *the sample* is prepared to sacrifice so that the change in health state leaves the sample's average overall utility unchanged between the two options.

This is implemented by assuming, as in TTO, that for each profile made up from living in state $\boldsymbol{x}_j$ for 10 years, there is a number of years (t <10) in full health which generates the same level of utility as this. So from equations 3 and 4, the probability of choosing the profile describing living in full health (11111) for t years is equal to the probability of choosing a profile describing living in a particular health state $\boldsymbol{x}_j$ for 10 years.[b]

$$\frac{exp(\hat{\alpha}+\hat{\beta}_2 t)}{exp(\hat{\alpha}+\hat{\beta}_2)+exp(\hat{\alpha}+\hat{\beta}_2 10+\hat{\beta}_3'\boldsymbol{x}_j\cdot 10)}=\frac{exp(\hat{\alpha}+\hat{\beta}_2 10+\hat{\alpha}+\hat{\beta}_3'\boldsymbol{x}_j\cdot 10)}{exp(\hat{\alpha}+\hat{\beta}_2 t)+exp(\hat{\alpha}+\hat{\beta}_2 10+\hat{\beta}_3'\boldsymbol{x}_j\cdot 10)} \tag{5}$$

The objective here is to derive the mean utility value of state $\boldsymbol{x}_j$ based on DCE that corresponds to a 10-year TTO value, which is t/10. Equation 5 can be solved so that this value is expressed as a function of the regression estimates:

$$\hat{V}_j^{DCE}=\frac{t}{10}=1+\frac{\hat{\beta}_3{}'}{\hat{\beta}_2}\boldsymbol{x}_j \tag{6}$$

[b] Since when $\boldsymbol{x}$ represents full health it is simply a vector of zero's. Alternatively effects coding could be applied to the data and a value for $\boldsymbol{\beta}'_3\boldsymbol{x}_j$.10 (where $\boldsymbol{x}_j$=full health) could be applied, but as the estimation relies on the difference between levels for each attribute, it can be proven that the results for the rescaled estimates from equation 6 are identical.

Thus, the sample mean DCE_{TTO} value for state $\boldsymbol{x}_j$ can be calculated from the coefficients of the conditional logit model.

4.3 Exclusion of respondents

There is a strong argument to exclude results from both the TTO and DCE_{TTO} where individuals have failed to understand or pay attention to the elicitation process as their responses will not necessarily represent their preferences (Devlin et al. 2003). Likewise including individuals that do not display compensatory behaviour violates the underpinning assumptions of consumer theory on which many choice based methods are based (Scott 2002). In the context of a web survey, engagement of respondents is expected to be more problematic than interview-based administration. However, excluding respondents is also problematic, not only as the statistical efficiency is reduced, but also because tests to identify 'irrational' respondents and 'lexicographic' preferences are deficient (Lancsar and Louviere 2006); thus, valid preferences may mistakenly be removed. When the objective of a study is to generate 'representative' preferences of society, such exclusions might compromise the results.

A series of criteria for detecting values that are deemed potentially problematic are employed. It is acknowledged that these criteria are imprecise and subjective, but use the results to generate a sample that appear to have no data problems for each exercise which is used to derive preliminary results. As suggested by Lancsar and Louviere (2006) the impact of including respondents that appeared to have increasingly more data problems is then examined in terms of their influence on model estimates.

For the TTO, respondents were potentially excluded if they: (i) answered preliminary (dominated) question incorrectly, (ii) had the values for all 5 health states were the same, (iii) had a given number of responses or more at 0.5 (which is the starting point for states considered better than dead, and so if the respondent wanted to complete the task quickly, 0.5 would be chosen), (iv) had a given number or more pairwise logical inconsistencies (as defined by Devlin et al., 2003) were found, and (v) had a given number or more health states valued worse than dead (whether a health state is worse than dead is the first choice in the task, so could easily be chosen by a respondent unengaged or not understanding the task) . To determine the exact criteria to use, we employed the technique illustrated by Devlin and colleagues where the impact of modifying the criteria was used to find the largest sample where values did not systematic differ between groups (see Devlin et al., 2003, for further details). This determined that respondents with greater than one pairwise logical inconsistency and 4 or more values worse than dead out of the 5 TTO tasks were excluded.

For the DCE_{TTO}, respondents were excluded if they: (i) answered both dominated questions incorrectly, (ii) had all 8 responses on the same side (all choices the left profile or all the right), (iii) had lexicographic preferences (where in all 8 tasks, individuals chose the profile with the best level of one attribute) and (iv) gave too little time to consider the task (defined as 8 seconds).

Consequently 5 separate samples are analysed: all_TTO refers to respondents that completed the TTO; noproblems_TTO refers to a subset of all_TTO that completed the TTO and did not have any potential data problems; all_DCE and noproblems_DCE are similarly also developed, and finally no problems_TTODCE refers to completers of both the TTO and DCE demonstrating no potential data problems in either exercise.

All analysis was performed in Matlab using Train's code (Train 2003) and SAS 9.1.

4.4 Model comparisons

The observed TTO values for the 48 health states were first compared to the predicted values from the TTO and DCE_{TTO} models. While observed TTO values are not a gold standard for comparing to the DCE_{TTO}, they provide an interesting comparison between the approaches. Previous studies have in the past used a variety of different tests to identify levels of correlation and agreement. We use a battery of tests that include the Pearson correlation coefficient and the intra-class correlation co-efficient (ICC) for comparisons between observed and predicted values, and for comparison between the mean observed and predicted values by health state, the root mean square difference (RMSD), the mean absolute difference (MAD), and the proportion of health state values predicted to within ±0.05 and ±0.1 of the observed mean of TTO values.

The difference between the 243 estimated health states of the EQ-5D are then compared *within* each elicitation technique based on the inclusion of respondents with potential data problems. A similar battery of tests are used as above, but since standard errors are estimated using shared covariances, the number of values with differences that are statistically significant ($p<0.05$) using paired t-tests is also reported. Systematic differences between values are observed using Bland-Altman plots (Bland and Altman 1986).

Comparisons between the final DCE_{TTO} and TTO models are then made via graphical means. Finally, the results of the self report responses are compared using chi-squared tests, and the time taken to complete each of the tasks reported.

5. Results

5.1 The sample

A sample of 4189 members of the market research panel was initially invited by email to participate in the survey. Of these 1400 (33%) consented to begin the survey and 1157 (83% of those that consented) completed the survey of both the TTO and DCE_{TTO}. In total, of the 1355 respondents that started the TTO exercise, 1175 (87%) completed all tasks. Some 10% of the respondents that failed to complete the TTO exercise dropped out at the first or second task, though this was less in respondents that had already completed the DCE_{TTO} exercise. Of the 1275 respondents that started the DCE_{TTO} exercise, 1220 (96%) completed all the tasks (Figure 1).

Figure 1 also describes the potential data problems identified in the sample. For the TTO overall, 62% of respondents had 1 or more pairwise inconsistencies within their 5 valuations. Similarly high rates have been found in previous studies (Lamers et al. 2006). Since the first question of the TTO task asks whether a state is worse than dead, the high number of values worse than dead is not surprising in respondents that were not engaged or did not understand the task. For the DCE_{TTO} 412 (34%) of respondents had potentially lexicographic preferences, mostly related to the life years attribute (e.g. choosing the profile with the longest life no matter the other attributes).

In total 537 (46%) of the TTO values and 527 (43%) of the DCE_{TTO} values were flagged as potentially problematic leaving 638 and 693 respondents in the samples noproblems_TTO and noproblems_DCE respectively. It should be noted that since the criteria for problems differ between techniques, these numbers should not be directly compared. Of the 1157 respondents that completed both the DCE_{TTO} and TTO exercises, 363 (31%) were defined as having no problems in either exercise (noproblems_TTODCE).

The characteristics of respondents are shown in Table 1. Respondents to the survey were older than non respondents (56.68 vs 48.20, $p<0.001$). Those respondents completing the tasks were older than those that did not complete the task (56.12 vs 59.36, $p<0.001$). The respondents with potential data problems tended to be younger than those with no potential data problems, but this was not statistically significant. The influence of differences in age likely impacted the education, income and marital status of respondents in each group. The EQ-5D profiles of respondents were similar to a previous study in the Canadian population (Johnson and Pickard 2000).

5.2 TTO model results

The coefficients for the random effect models are shown in Table 2. With the exception of the usual activities attribute, all the coefficients from the TTO analysis were logically consistent. In first sample (noproblems_TTODCE), levels 2 and 3 of the usual activities attribute were disordered but were not significantly different from each other. The model had an R-square (square of the Pearson correlation coefficient) above 0.40, similar to previous TTO EQ-5D studies and relatively good predictive performance (RMSD less than 0.07). None of the analyses suffered from multicollinearity. However, consistent with previous studies (e.g. Dolan 1997; Shaw et al. 2005) residuals were only approximately normally distributed, and the RESET test suggested the presence of misspecification due to omitted variables or incorrect functional form.

Adding respondents who had data problems with the DCE (but not the TTO), defined as the sample noproblems_TTO show there is high agreement between values with an ICC of 0.994 and no values with a difference greater than 0.1. Levels 2 and 3 of the usual activities attribute become logically consistent. However, the inclusion of respondents with potential data problems (all_TTO) has a large impact on the coefficient estimates, notably the constant which increases from approximately 0.1 to close to 0.5. The Bland Altman plot (Figure 2a) and comparison statistics strongly suggest that the values obtained from the sample all_TTO are systematically different from the values obtained from noproblems samples. It was consequently decided to use noproblems_TTO as the final TTO values as these represent the largest sample of respondents that appear to have understood and engaged with the

TTO task. The values from this sample (1 to -0.384) are within the range of values estimated in EQ-5D TTO studies from other countries (Szende et al 2007).

5.3 DCE$_{TTO}$ results

The results from the random effects conditional logit model are presented in Table 3. Each of the samples has attribute coefficients with the expected sign so that on average respondents preferred to live in longer health profiles and in less severe levels of each health status attribute and are consistently ordered. The coefficient for the constant term is not significant in any model, suggesting there is no specification error in the analyses (Scott 2001). Nearly all the coefficient values are significant at the conventional significance levels. The inclusion of $\boldsymbol{\beta}_1$ terms did lead to multicollinearity with correlation coefficients between each corresponding attribute level in $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_3$ over 0.7[c].

The re-anchored coefficients are shown in Table 2. While the estimated values have relatively high correlations, they poorly predict the TTO observed values. The ICCs are close to 0.6 but over half of the 48 health states have a difference that is greater than 0.1. However, Table 4 shows there is more agreement in health states with few observed values considered worse than dead (WTD). For example while the overall MAD between model 6 of the DCE$_{TTO}$ and the TTO is 0.141, if this is separated into health states with 0% values WTD, 0 to 10% values WTD, 10% to 50% values WTD and over 50% values WTD, the MAD varies from -0.029 to 0.067 to 0.165 to 0.434. Since the values considered WTD are derived using arbitrary transformations, this suggests there may actually be relatively good agreement between DCE$_{TTO}$ and the TTO values.

The differences between estimated values from each of the DCE$_{TTO}$ samples are also summarized in Table 2. In comparison to values from the noproblems_TTODCE sample, an ICC of 0.981 suggests there is little difference in values estimated using the sample noproblems_DCE when the respondents who were deemed to have TTO data problems are included. While 70 of the 243 values are different by greater than 0.1, only 1 of these differences is statistically significant. When the respondents with DCE data problems are also added (all_DCE), in comparison to the noproblems_DCETTO sample there is even more agreement between estimated values. The ICC improves to 0.991 and only 17 health states

[c] Mark and Swait (2008) state as a rule of thumb that problems of multicollinearity are likely to occur if any of the correlations between any of the independent variables are greater than 0.7

have a difference greater than 0.01, none of which are statistically significant. This suggests there is little difference in preferences between the sample of respondents with no data problems and the sample that includes respondents with any data problems. The Bland Altman plots in Figure 2 also suggest there are no systematic differences between values. The all_DCE sample is used as for the final values since the inclusion of more respondents reduces the variance in coefficients and there is no reason to exclude respondents with apparent DCE problems. The estimated for range from between 1 (health state 11111) and -1.133 (health state 33333).

5.5 Comparison of TTO and DCE_{TTO}

The estimated values for the 243 health states for the EQ-5D based on the final TTO and DCE_{TTO} samples are shown in Table 4 and Figure 3 and a full list is in the appendix. There is clear divergence at the very mild health states, where the DCE_{TTO} estimates values close to 1 (the highest value apart from full health is 0.956 for state 21111) while the TTO has a gap from full health to impaired health (the highest value is 0.807 for state 21111). There is also divergence in the lower health states. The DCE_{TTO} estimates 55 states to be worse than dead, while the TTO estimates only 15.

The mean time taken to complete the whole survey was 22 minutes (IQR 14-26). As expected respondents took more time in completing the first question of each exercise (over 2.5 minutes for the TTO and just over 1 minute for the DCE), than in subsequent questions (average of just under 2 minutes for the TTO and just over 30 seconds for the DCE). To complete the 5 TTO valuations, respondents took 9.5 minutes to complete the 5 TTO valuations. This compared to under 4 minutes to complete the 8 DCE_{TTO} tasks. Times did not vary significantly between the valuation sample and respondents with data problems.

In the valuation sample, there was little difference in self-reported difficulty in understanding or answering the two exercises with under 15% of respondents finding each exercise fairly or very difficult to understand, and 50% fairly or very difficult to answer. Respondents with data problems found the tasks harder to understand, but simpler to answer than the valuation sample.

6. Discussion

This paper presents a new method for estimating health state values through the use of a DCE. The values appear robust, with estimated coefficients that are statistically significant, logically consistent and with expected signs. The TTO values from the study cannot be considered a 'gold standard' with which to compare the values generated from DCE_{TTO} since they are derived using different economic theories, each requiring different assumptions for econometric modelling. They do however provide a basis for comparison, and give context to a discussion to the wider merits and implications of using DCE_{TTO} as an alternative to the TTO.

A principle finding from this study is that in contrast to the TTO, the inclusion of respondents that may not have understood, not engaged or were 'irrational' in the DCE_{TTO} had little influence on the results. Lancsar and Louviere (2006) suggest that the random component in RUT, often referred to as unobservable, can be interpreted to capture errors made by 'irrational' respondents. It is also possible that the DCE_{TTO} was cognitively easier for respondents than the TTO leading to fewer data problems. Given the design, we cannot establish whether one technique was necessarily cognitively easier to another. However, the self-report results suggested that the DCE_{TTO} was at least not more cognitively difficult than the TTO, there were fewer respondents that did not complete the task, and the DCE_{TTO} took less time to complete than the TTO.

The implication is that using the DCE_{TTO} can potentially reduce the bias associated with excluding certain respondents. This is particularly important in valuation studies where the objective is to estimate representative values from the general population. In TTO studies, researchers have to decide which respondents with data problems have 'crucially failed to understand the task'. This is very difficult to determine, largely subjective and based on deficient tests – and in our study had a large impact on estimated values. While our study likely magnifies the number of respondents with potential data problems by using a web survey instead of an interview, previous interview-based TTO studies have typically excluded some respondents from their final valuations suggesting the issue is still present.

Buckingham and Devlin (2006) have recently provided a theoretical underpinning for the TTO drawing on Hicks'utility theory. However, individuals are asked to make choices between certain outcomes, and can only trade the number of years in the task. That DCEs are rooted in RUT is a benefit to their use in health state valuations. In the DCE_{TTO}, individuals are asked to trade between attributes describing both levels of health status and life years. Ranking data in the past have been exploded into a series of pairwise choices to reflect DCE results (e.g. Salomon 2003, McCabe et al. 2006), but are less consistent with RUT in comparison to DCEs (Louviere et al. 2000). Ranking data also requires assumptions that the ordering of a pair of health states does not depend on the other states being considered (independent of irrelevant alternatives), which is unlikely to be satisfied.

A methodological advantage to the DCE_{TTO} is that health states can be valued worse than dead without altering the task, and without transformations of resulting values, as is done in the conventional TTO. Lamers (2007) demonstrate the subjectivity of the different approaches used in such transformations and show how they can impact average values substantially. This has led to the development of new approaches in the TTO that are uniform across states better and worse than dead, such as the "lead time" TTO (Devlin et al. 2009). In the DCE_{TTO}, health states can be valued worse than dead indirectly. The model results derive the relative preference between each health state on the latent scale. The latent scale is then anchored on the health utility scale, essentially finding the point on the scale where, on average, the values become worse than dead. Given the unfamiliarity most individuals have with health states potentially worse than dead, such an indirect approach has significant appeal. However negative values obtained from DCE_{TTO} are in essence extrapolated to the negative range based on data in the positive range. The DCE_{TTO} approach could be extended to include a lead time in each profile.

The modelling of both elicitation methods used in this paper has assumed that the utility function for additional life years is linear in time. Tsuchiya and Dolan (2005) find in a review of existing TTO studies that the assumption of constant proportional time trade off holds on the aggregate level, but is violated at the individual level. Testing this assumption in the TTO requires experiments to be repeated with different survival baselines. In contrast, the DCE_{TTO} enables this assumption to be explored internally without repeating the experiment by modelling the life-years terms as categorical variables.

Some limitations of the study surround the experimental design of the DCE_{TTO}, and the representativeness of the invited sample. Experimental design issues for the DCE_{TTO} are more complex than the TTO as the pairing of profiles can inadvertently lead to covariances between attributes, and the valuation space, with 6 attributes, is much larger. The implausible attribute-level combinations contained in the EQ-5D led to attributes with moderate correlations in the design. Further interactions between health attribute levels, which are typically included in EQ-5D valuations, were also not accounted for. In this respect, the experimental design used in this study could most certainly be improved since while covariances between attributes were small, they could influence the estimates. The results of this study could be used to estimate Bayesian optimal designs in a future study (Bliemer et al. 2008). Furthermore, while respondents were broadly representative of the Canadian general population in terms of age, gender and education, there may be concerns since they were members of a market research panel. The practical advantage to this approach to recruitment was that rapid and inexpensive valuations can be obtained without any potential interviewer bias. However, the confirmation of these results in a sample of respondents not from a market research panel, perhaps in the form of an interview rather than a web survey, is desirable.

The resources required to undertake valuation studies depends on the number of respondents recruited that complete the tasks producing usable values, the number of tasks each respondent is asked to complete, the time taken to do this, mode of administration and the experimental design. The findings can only provide a rudimentary insight into which elicitation technique would require the most resources to obtain similar precision in estimates. While the DCE_{TTO} produced values with larger variances, in overall terms, the resources required for the DCE_{TTO} were no greater, and probably less than that for the TTO; this was principally due to a higher percentage of recruited respondents producing usable values and less time being required to complete tasks.

In summary, this study presents a new method for health state valuation using a stand-alone DCE design that produces values anchored on 1 for full health and 0 for dead. The approach is able to take account of states worse than dead in a single task, and its results are less prone to bias from excluding respondents therefore providing more representative values. Further research on the potential advantages and limitations of this approach are necessary, and work to identify if this approach might facilitate valuations in diverse settings and population groups is required.

References

Belsley DA, Kuh E, and R. E Welsch. (1980). Regression diagnostics. Wiley New York.

Bliemer MCJ, Rose JM, and Hess S. (2008). Approximation of bayesian efficiency in experimental choice designs. Journal of Choice Modelling 1, no. 1: 98.

Brazier J, Ratcliffe J, Salomon J, and Tsuchiya A. (2007). Measuring and Valuing Health Benefits for Economic Evaluation. 1st ed. Oxford University Press.

Brazier J., Deverill M, Green C, Harper R, and Booth A. (1999). A review of the use of health status measures. Economic Evaluation Health Technology Assessment 3, no. 9.

Brazier J, Rowan D, Yang Y, and Tsuchiya A. (2009). Using Rank and Discrete Choice Data to Estimate Health State Utility Values on the QALY Scale. HEDS Discussion paper 09/10. Available http://www.shef.ac.uk/scharr/sections/heds/discussion.html. Accessed Sept 2009

Brooks R. (1996). EuroQol: the current state of play. Health Policy. 37, no. 1: 53-72.

Bryan S, and Dolan P. (2004). Discrete choice experiments in health economics. The European Journal of Health Economics 5, no. 3: 199-202.

Buckingham K and Devlin N. (2006). A theoretical framework for TTO valuations of health. Health Economics 15, no. 10: 1149-1154.

Burr JM, Kilonzo M, Vale L, and Ryan M. (2007). Developing a preference-based Glaucoma Utility Index using a discrete choice experiment. Optometry and Vision Science. 84, no. 8: 797-808.

Coast J, Flynn TN, Natarajan L, Sproston K, Lewis J, Louviere J, and Peters TJ. (2008). Valuing the ICECAP capability index for older people. Social Science & Medicine 67, no. 5: 874-82.

Craig BM, Busschbach JJV, and Salomon JA. (2009). Keep it simple: ranking health states yields values similar to cardinal measurement approaches. Journal of Clinical Epidemiology 62, no. 3: 296-305.

Devlin N, Tsuchiya, A, Buckingham, K, and Tilling, C. (2009) A Uniform Time trade Off Method for States Worse than Dead: Feasibility Study of the 'Lead Time' Approach. City University London Discussion paper 09/08. Available http://www.city.ac.uk/economics/dps/discussion_papers/0908.pdf. Accessed Sept 2009

Devlin NJ, 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, no. 7: 529-544.

Dolan P. (1997). Modeling valuations for EuroQol health states. Medical Care 35, no. 11: 1095-1108.

Flynn TN, Louviere JJ, Marley AA, Coast J, and Peters TJ. (2008). Rescaling quality of life values from discrete choice experiments for use as QALYs: a cautionary tale. Population Health Metrics 6: 6.

Gudex C. (1994). Time Trade-off User Manual: Props and Self-completion Methods. Centre for Health Economics Discussion paper. http://www.york.ac.uk/inst/che/pdf/op20.pdf

Hakim Z, and Pathak DS. (1999). Modelling the EuroQol data: a comparison of discrete choice conjoint and conditional preference modelling. Health Economics 8, no. 2: 103-116.

Johnson JA, and Pickard AS. (2000). Comparison of the EQ-5D and SF-12 health surveys in a general population survey in Alberta, Canada. Medical Care 38, no. 1: 115-121.

Kuhfeld WF. (2009). Orthogonal arrays. Available http://support.sas.com/techsup/technote/ts723.html. Accessed Sept 2009

Kuhfeld WF. (2005). Marketing Research Methods in SAS: Experimental Design, Choice, Conjoint, and Graphical Techniques. Technical report, SAS Institute. Available at http://support.sas.com/techsup/technote/ts723.html. Accessed Sept 2009

Lamers LM, McDonnell J, Stalmeier PFM, Krabbe PFM, and Busschbach JJV. (2006). The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Economics 15, no. 10: 1121-1132.

Lamers LM, Stalmeier PFM, Krabbe PFM, and Busschbach JJV. (2006). Inconsistencies in TTO and VAS values for EQ-5D health states. Medical Decision Making: An International Journal of the Society for Medical Decision Making 26, no. 2: 173-181.
Lamers LM. (2007). The transformation of utilities for health states worse than death - Consequences for the estimation of EQ-SD value sets. Medical care 45, no. 3: 238-244.
Lancsar E, and Donaldson C. (2005). Discrete choice experiments in health economics. The European Journal of Health Economics 6, no. 4: 314–316.
Lancsar E, and Louviere J. (2006). Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Economics 15, no. 8: 797-811.
Louviere JJ, and Woodworth G. (1983). Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. Journal of marketing research 20, no. 4: 350–367.
Louviere JJ, Hensher DA, and Swait JD. (2000). Stated Choice Methods: Analysis and Applications. Cambridge University Press.
Mark T, and Swait J. (2008). Using stated preference and revealed preference data fusion modelling in health care. In Using Discrete Choice Experiments to Value Health and Health Care. Springer.
Bland MJ, and Altman DG. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 327, no. 8476: 307–310.
McCabe C, Brazier J, Gilks P, Tsuchiya A, Roberts J, O'Hagan A, and Stevens K. (2006). Using rank data to estimate health state utility models. Journal of Health Economics 25, no. 3: 418-31.
Mcfadden D. (1974). Conditional logit analysis of qualitative choice behavior. In Frontiers in Econometrics. P. Zarembka, Ed. Academic Press, New York.
Oehlert GW. (1992). A note on the delta method. American Statistician: 27–29.
Ramsey J B. (1969). Tests for specification errors in classical linear least-squares regression analysis. Journal of the Royal Statistical Society. 31, no. 2: 350–371.
Ratcliffe J, Brazier J, Tsuchiya A, Symonds T, and Brown M. (2009). Using DCE and ranking data to estimate cardinal values for health states for deriving a preference-based single index from the sexual quality of life questionnaire. Health Economics 18, no 11: 1261-76.
Ryan M, Netten A, Skatun D, and Smith P. (2006). Using discrete choice experiments to estimate a preference-based measure of outcome--An application to social care for older people. Journal of Health Economics 25, no. 5: 927-944.
Salomon J. (2003). Reconsidering the use of rankings in the valuation of health states: a model for estimating cardinal values from ordinal data. Population Health Metrics 1, no. 1: 12.
Scott A. (2001). Eliciting GPs' preferences for pecuniary and non-pecuniary job characteristics. Journal of Health Economics 20, no. 3: 329–347.
Scott A. (2002). Identifying and analysing dominant preferences in discrete choice experiments: an application in health care. Journal of Economic Psychology 23, no. 3: 383–398.
Shaw JW, Johnson JA, and Coons SJ. (2005). US Valuation of the EQ-5D Health States: Development and Testing of the D1 Valuation Model. Medical Care 43, no. 3: 203.
Szende AM, Oppe M, and Devlin N. (2007). EQ-5D Value Sets: Inventory, Comparative Review and User Guide. Series: EuroQol Group Monographs 2.
Torrance GW. (1986). Measurement of health state utilities for economic appraisal. Journal of Health Economics 5, no. 1: 1-30.
Train K. (2003). Discrete choice methods with simulation. Cambridge University Press.
Tsuchiya A, and Dolan P. (2005). The QALY model and individual preferences for health states and health profiles over time: a systematic review of the literature. Medical Decision Making. 25, no. 4: 460-467.

Viney R, Savage E, King M, and Hossain I. (2007). Using Choice Experiments to Estimate QALYs: An Application to Prostate Cancer. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=994063. Accessed Sept 2009

Table 1. Characteristics of the Sample

Characteristic	1. Completers with no data problems (n=363)	2. Completers with some data problems (n=794)	3. Non completers (n=243)	4. Non respondents (n=2789)	P Value†		
					1 vs 2	1+2 vs 3	1+2+3 vs 4
Age, mean (SE)	56.87 (0.70)	55.78 (0.48)	59.36 (0.80)	48.20 (0.31)	0.199	<0.001	<0.001
Sample range	18-83	18-99	27-83	18-94			
Sex, % (n)							
Male	161 (44)	384 (48)	117 (48)	1338 (48)	0.205	<0.001	0.673
Female	202 (56)	410 (52)	126 (52)	1451 (52)			
Highest level of education, n (%)							
Primary school or less	0 (0)	3 (0)	1 (0)	31 (1)	0.711	<0.001	<0.001
High School	78 (21)	171 (22)	56 (23)	749 (27)			
Community college	136 (37)	309 (39)	102 (42)	1105 (40)			
Undergraduate degree	109 (30)	236 (30)	62 (26)	669 (24)			
Graduate degree	40 (11)	75 (9)	22 (9)	235 (8)			
Income CAD, n (%)							
$10,000 or less	1 (0)	15 (2)	7 (3)	33 (1)	0.124	<0.001	<0.001
$10,000-$20,000	19 (5)	37 (5)	7 (3)	269 (10)			
$20,000-$30,000	32 (9)	61 (8)	25 (10)	325 (12)			
$30,000-$40,000	41 (11)	99 (12)	40 (16)	455 (16)			
$40,000-$60,000	87 (24)	221 (28)	60 (25)	729 (26)			
$60,000-$80,000	63 (17)	152 (19)	46 (19)	501 (18)			
$80,000-$100,000	56 (15)	99 (12)	32 (13)	253 (9)			
$100,000 or more	64 (18)	110 (14)	26 (11)	224 (8)			
Marital status, n (%)							
Married	218 (60)	444 (56)	135 (56)	1275 (46)	0.667	<0.001	<0.001
Living with partner	24 (7)	69 (9)	20 (8)	271 (10)			
Widowed	24 (7)	53 (7)	23 (9)	135 (5)			
Divorced	40 (11)	77 (10)	23 (9)	262 (9)			
Separated	11 (3)	31 (4)	11 (5)	140 (5)			
Single	44 (12)	117 (15)	30 (12)	696 (25)			
Not reported	2 (1)	3 (0)	1 (0)	10 (0)			
EQ-5D dimension, n (%)							
Mobility							
Problems	92 (25)	197 (25)	65 (27)	-	0.846	0.564	-
No problems	271 (75)	597 (75)	178 (73)	-			
Self-care							
Problems	15 (4)	28 (4)	14 (6)	-	0.613	0.143	-
No problems	348 (96)	766 (96)	229 (94)	-			
Usual activities							
Problems	96 (26)	161 (20)	52 (21)	-	0.019	0.781	-
No problems	267 (74)	633 (80)	191 (79)	-			
Pain/discomfort							
Problems	197 (54)	425 (54)	139 (57)	-	0.814	0.327	-
No problems	166 (46)	369 (46)	104 (43)	-			
Anxiety/depression							
Problems	85 (24)	210 (26)	69 (28)	-	0.397	0.389	-
No problems	268 (76)	584 (74)	174 (72)	-			
No problems in any dimension	141 (40)	299 (38)	67 (28)	-	0.700	0.002	-
EQ-5D UK index, mean (SE)	0.80 (0.01)	0.80 (0.01)	0.80 (0.01)	-	0.838	0.823	-
EQ-5D US index, mean (SE)	0.85 (0.01)	0.85 (0.01)	0.84 (0.01)	-	0.880	0.563	-

† T-test for continuous data, chi-square test for categorical data

Table 2: Parameter Estimates from modelled TTO and rescaled DCE_{TTO} models (on the health utility scale)

Variable	TTO: Random effects model Estimate (SE)			DCE_{TTO}: Re-anchored estimates Estimate (SE†)		
	Model 1 *Group I (n=363)*	**Model 2** *Group II (n=638)*	**Model 3** *Group III (n=1177)*	**Model 4** *Group I (n=363)*	**Model 5** *Group II (n=693)*	**Model 6** *Group III (n=1220)*
Mobility level 2	0.026 (0.026)	0.023 (0.020)	0.003 (0.018)	0.092 (0.056)	0.085 (0.042)**	0.050 (0.031)
Mobility level 3	0.326 (0.031)**	0.309 (0.024)**	0.172 (0.020)**	0.599 (0.064)**	0.596 (0.049)**	0.563 (0.036)**
Self-care level 2	0.070 (0.026)*	0.078 (0.021)**	0.105 (0.018)**	0.056 (0.066)	0.065 (0.049)	0.085 (0.036)**
Self-care level 3	0.230 (0.027)**	0.235 (0.021)**	0.198 (0.018)**	0.351 (0.065)**	0.398 (0.049)**	0.393 (0.036)**
Usual Activities level 2	0.107 (0.028)**	0.097 (0.022)**	0.112 (0.019)**	0.029 (0.071)	0.103 (0.053)	0.089 (0.039)**
Usual Activities level 3	0.101 (0.032)*	0.124 (0.025)**	0.087 (0.021)**	0.172 (0.070)**	0.243 (0.053)**	0.238 (0.039)**
Pain /discomfort level 2	0.061 (0.027)*	0.036 (0.021)	0.050 (0.018)*	0.110 (0.063)	0.115 (0.048)**	0.095 (0.036)**
Pain /discomfort level 3	0.334 (0.026)**	0.315 (0.020)**	0.203 (0.017)**	0.527 (0.064)**	0.501 (0.049)**	0.447 (0.036)**
Anxiety /depression level 2	0.052 (0.027)	0.070 (0.022)**	0.036 (0.018)	0.094 (0.061)	0.115 (0.046)**	0.104 (0.034)**
Anxiety /depression level 3	0.299 (0.027)**	0.294 (0.021)**	0.214 (0.018)**	0.372 (0.064)**	0.395 (0.049)**	0.384 (0.036)**
Constant	0.118 (0.030)**	0.107 (0.024)**	0.486 (0.022)**	-	-	-
Number of observations	1815	3190	5875	2178	4158	7320
Predictive performance (48 TTO health states)						
Correlation	0.677	0.653	0.411	0.671	0.671[a]	0.649[a]
ICC	0.854	0.844	0.454	0.594	0.589[a]	0.601[a]
MAD	-0.004	-0.006	0.007	0.106	0.197[a]	-0.141[a]
RMSD	0.073	0.052	0.063	0.148	0.212[a]	0.163[a]
n >|0.05|	30	25	27	36	40[a]	35[a]
n >|0.10|	14	7	9	26	32[a]	27[a]
Model comparisons (with values from 243 health states from group 1)						
Correlation	ref	0.997	0.966	ref	0.995	0.992
ICC	ref	0.994	0.612	ref	0.981	0.991
MAD	ref	-0.019	0.225	ref	0.072	0.016
RMSD	ref	0.024	0.226	ref	0.074	0.045
n >|0.05|	ref	17	229	ref	168	103
n >|0.10|	ref	0	213	ref	70	17
n diff stat sig	ref	0	200	ref	1	0

*p<0.05

**p<0.001

a - compared to TTO values from 638 respondents in model 2

† SEs calculated using the delta method (Oehlert 1992), average SEs for each attribute level presented and used in paired t-tests

Table 3: Parameter estimates from the DCE_{TTO} models (on the latent utility scale)

Variable	Parameter	Conditional logit, estimate (SE)		
		Model 4 *Group I (n=363)*	**Model 5** *Group II (n=693)*	**Model 6** *Group III (n=1220)*
Life years	Mean coefficient S.D of coefficient	0.435 (0.028)**	0.420 (0.020)**	0.435 (0.015)**
Mobility level 2 x Life years	Mean coefficient S.D of coefficient	-0.040 (0.013)**	-0.036 (0.009)**	-0.022 (0.007)**
Mobility level 3 x Life years	Mean coefficient S.D of coefficient	-0.261 (0.017)**	-0.250 (0.012)**	-0.245 (0.009)**
Self-care level 2 x Life years	Mean coefficient S.D of coefficient	-0.025 (0.014)	-0.028 (0.010)**	-0.037 (0.007)**
Self-care level 3 x Life years	Mean coefficient S.D of coefficient	-0.153 (0.014)**	-0.167 (0.010)**	-0.171 (0.008)**
Usual Activities level 2 x Life years	Mean coefficient S.D of coefficient	-0.013 (0.015)	-0.043 (0.011)**	-0.039 (0.008)**
Usual Activities level 3 x Life years	Mean coefficient S.D of coefficient	-0.075 (0.015)**	-0.102 (0.011)**	-0.104 (0.008)**
Pain /discomfort level 2 x Life years	Mean coefficient S.D of coefficient	-0.048 (0.014)**	-0.049 (0.010)**	-0.041 (0.008)**
Pain /discomfort level 3 x Life years	Mean coefficient S.D of coefficient	-0.229 (0.015)**	-0.210 (0.011)**	-0.194 (0.008)**
Anxiety /depression level 2 x Life years	Mean coefficient S.D of coefficient	-0.041 (0.013)**	-0.048 (0.009)**	-0.045 (0.007)**
Anxiety /depression level 3 x Life years	Mean coefficient S.D of coefficient	-0.162 (0.015)**	-0.166 (0.011)**	-0.167 (0.008)**
Constant	Mean coefficient	0.080 (0.054)	0.058 (0.038)	0.041 (0.030)**
Number of observations		2178	4158	7320
Log likelihood		-1103	-2144	-3734
P (correct)		75.1	74.3	74.2

* significant from zero p<0.05
** significant from zero p<0.01

Table 4: Observed TTO and predicted TTO (model 2) and DCE$_{TTO}$ (model 6) values

Health state	1. Observed TTO mean (SE)	1. Observed TTO % WTD	2. Predicted TTO, mean (SE)	3. Predicted DCE$_{TTO}$ mean (SE)	Mean Difference 1 vs 2	Mean Difference 1 vs 3	Mean Difference 2 vs 3
21111	0.879 (0.023)	0%	0.870 (0.027)	0.950 (0.002)	0.009	-0.071	-0.080
12111	0.874 (0.021)	0%	0.815 (0.027)	0.915 (0.024)	0.059	-0.041	-0.100
11121	0.861 (0.024)	0%	0.856 (0.027)	0.905 (0.024)	0.005	-0.044	-0.049
11112	0.856 (0.027)	0%	0.823 (0.026)	0.896 (0.003)	0.033	-0.040	-0.073
21112	0.818 (0.033)	0%	0.801 (0.028)	0.847 (0.005)	0.017	-0.029	-0.046
21212	0.810 (0.027)	0%	0.704 (0.026)	0.758 (0.025)	0.106	0.052	-0.054
11211	0.807 (0.053)	2%	0.796 (0.028)	0.911 (0.024)	0.011	-0.104	-0.115
11222	0.766 (0.041)	3%	0.690 (0.028)	0.712 (0.034)	0.076	0.054	-0.022
22112	0.720 (0.037)	2%	0.722 (0.029)	0.762 (0.025)	-0.002	-0.042	-0.04
11312	0.708 (0.046)	2%	0.699 (0.031)	0.658 (0.026)	0.009	0.050	0.041
21321	0.672 (0.050)	2%	0.710 (0.033)	0.617 (0.035)	-0.038	0.055	0.093
23121	0.613 (0.043)	2%	0.598 (0.031)	0.462 (0.038)	0.015	0.151	0.136
31211	0.595 (0.041)	4%	0.487 (0.030)	0.348 (0.029)	0.108	0.247	0.139
11113	0.591 (0.047)	5%	0.599 (0.026)	0.616 (0.027)	-0.008	-0.025	-0.017
12321	0.589 (0.047)	2%	0.654 (0.032)	0.582 (0.043)	-0.065	0.007	0.072
22222	0.581 (0.055)	6%	0.589 (0.026)	0.578 (0.043)	-0.008	0.003	0.011
13311	0.560 (0.049)	5%	0.534 (0.033)	0.368 (0.039)	0.026	0.192	0.166
22131	0.554 (0.046)	4%	0.477 (0.031)	0.418 (0.038)	0.077	0.136	0.059
22123	0.548 (0.071)	12%	0.462 (0.032)	0.387 (0.045)	0.086	0.161	0.075
11131	0.547 (0.057)	7%	0.578 (0.028)	0.553 (0.028)	-0.031	-0.006	0.025
23311	0.506 (0.057)	9%	0.511 (0.032)	0.319 (0.040)	-0.005	0.187	0.192
12231	0.504 (0.061)	6%	0.403 (0.031)	0.379 (0.045)	0.101	0.125	0.024
23222	0.482 (0.062)	11%	0.432 (0.027)	0.269 (0.047)	0.050	0.213	0.163
13122	0.443 (0.068)	10%	0.551 (0.032)	0.408 (0.038)	-0.108	0.035	0.143
31221	0.409 (0.060)	16%	0.451 (0.030)	0.253 (0.038)	-0.042	0.156	0.198
21232	0.407 (0.058)	12%	0.389 (0.026)	0.310 (0.040)	0.018	0.097	0.079
12332	0.369 (0.059)	19%	0.306 (0.031)	0.126 (0.050)	0.063	0.243	0.180
22232	0.368 (0.068)	14%	0.311 (0.026)	0.225 (0.048)	0.057	0.143	0.086
13113	0.361 (0.078)	16%	0.363 (0.029)	0.223 (0.042)	-0.002	0.138	0.140
22313	0.331 (0.067)	20%	0.374 (0.030)	0.243 (0.048)	-0.043	0.088	0.131
32211	0.327 (0.071)	16%	0.409 (0.030)	0.263 (0.038)	-0.082	0.064	0.146
12213	0.326 (0.080)	24%	0.424 (0.030)	0.442 (0.044)	-0.098	-0.116	-0.018
13223	0.298 (0.076)	23%	0.230 (0.031)	0.039 (0.057)	0.068	0.259	0.191
11133	0.168 (0.076)	38%	0.284 (0.029)	0.169 (0.043)	-0.116	-0.001	0.115
32223	0.133 (0.088)	37%	0.078 (0.028)	-0.216 (0.055)	0.055	0.349	0.294
31323	0.124 (0.088)	35%	0.129 (0.027)	-0.280 (0.051)	-0.005	0.404	0.409
23232	0.085 (0.080)	37%	0.154 (0.025)	-0.083 (0.054)	-0.069	0.168	0.237
21233	0.084 (0.084)	40%	0.165 (0.029)	0.030 (0.052)	-0.081	0.054	0.135
23231	0.066 (0.087)	42%	0.223 (0.028)	0.021 (0.052)	-0.157	0.045	0.202
21333	0.065 (0.082)	32%	0.137 (0.030)	-0.119 (0.055)	-0.072	0.184	0.256
32313	0.015 (0.076)	38%	0.088 (0.025)	-0.270 (0.051)	-0.073	0.285	0.358
32323	0.004 (0.077)	47%	0.051 (0.026)	-0.365 (0.057)	-0.047	0.369	0.416
23233	-0.055 (0.076)	54%	-0.071 (0.026)	-0.363 (0.065)	0.016	0.308	0.292
32232	-0.057 (0.094)	52%	0.025 (0.030)	-0.288 (0.051)	-0.082	0.231	0.313
33323	-0.110 (0.074)	56%	-0.106 (0.023)	-0.674 (0.063)	-0.004	0.564	0.568
33213	-0.111 (0.072)	60%	-0.042 (0.029)	-0.429 (0.053)	-0.069	0.318	0.387
33332	-0.264 (0.056)	72%	-0.160 (0.026)	-0.746 (0.060)	-0.104	0.482	0.586
33333	-0.346 (0.020)	77%	-0.384 (0.017)	-1.026 (0.071)	0.038	0.680	0.642
MAD					-0.006	0.141	0.148

Table 5. Comparison in responses between methods

	No problems (n=363)			Data problems (n=794)			No problems vs data problems	
	TTO	**DCE$_{TTO}$**	**P-values†**	**TTO**	**DCE$_{TTO}$**	**P-values†**	**TTO P-values**	**DCE$_{TTO}$ P-values**
Difficulty in 'understanding', n (%)								
Very difficult	1 (0)	3 (1)	0.622	23 (3)	19 (2)	0.501	0.011	0.020
Fairly difficult	42 (12)	38 (10)		114 (14)	124 (16)			
Not very difficult	171 (47)	180 (50)		366 (46)	384 (48)			
Not at all difficult	149 (41)	139 (38)		289 (36)	264 (33)			
Missing	0 (0)	3 (1)		2 (0)	3 (0)			
Difficulty in 'answering', n (%)								
Very difficult	34 (9)	45 (12)	0.388	67 (8)	76 (10)	0.864	0.145	0.092
Fairly difficult	148 (41)	142 (39)		279 (35)	274 (35)			
Not very difficult	127 (35)	113 (31)		295 (37)	297 (37)			
Not at all difficult	54 (15)	62 (17)		153 (19)	147 (19)			
Missing	0 (0)	1 (0)		0 (0)	0 (0)			

† T-test for continuous data, chi-square test for categorical data

‡ Excluding rationality questions and individual questions beyond the first taking longer than 20 minutes, which were assumed to be time where the user was not considering the question but something else.

Figure 1. Description of survey responses

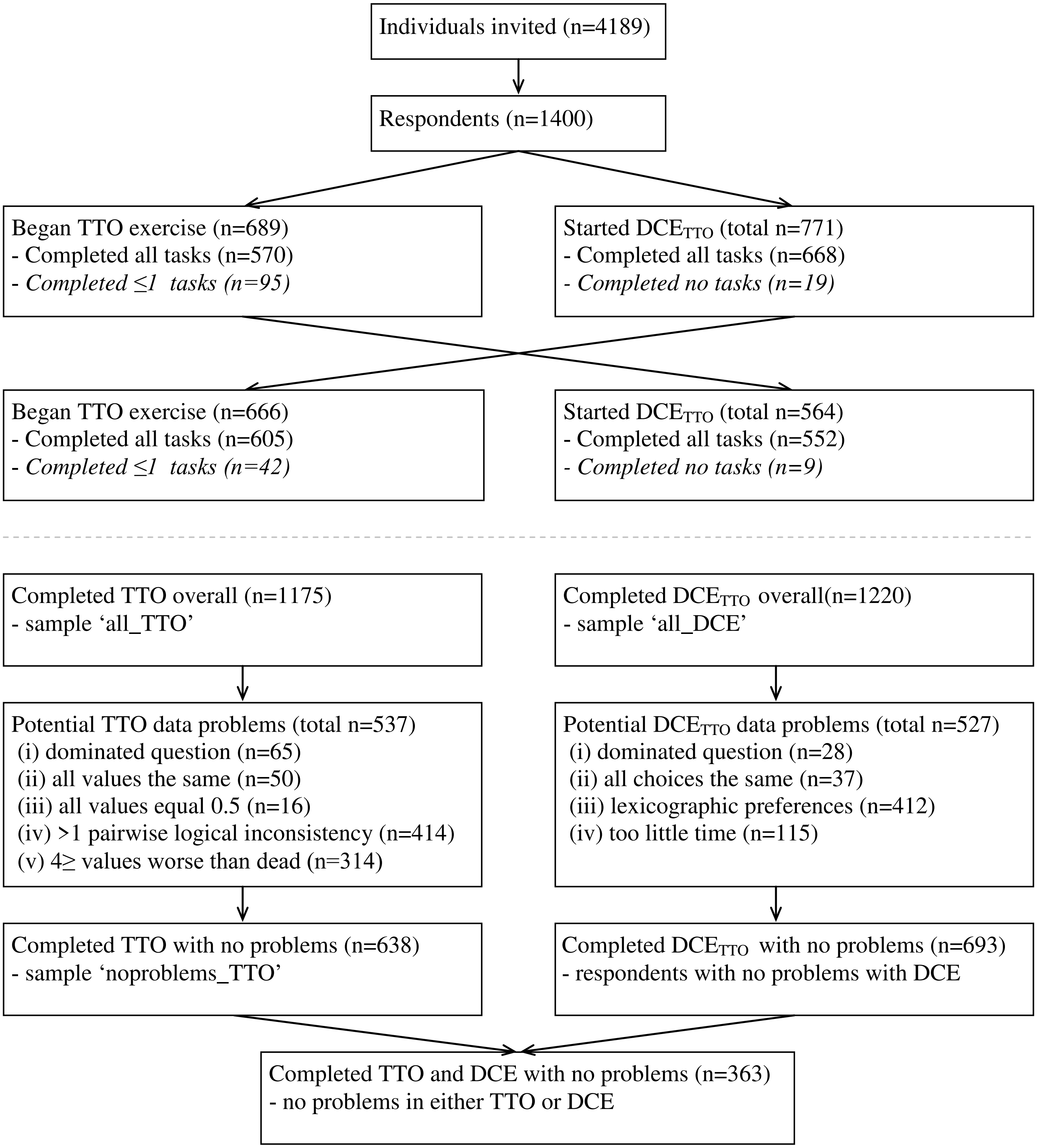
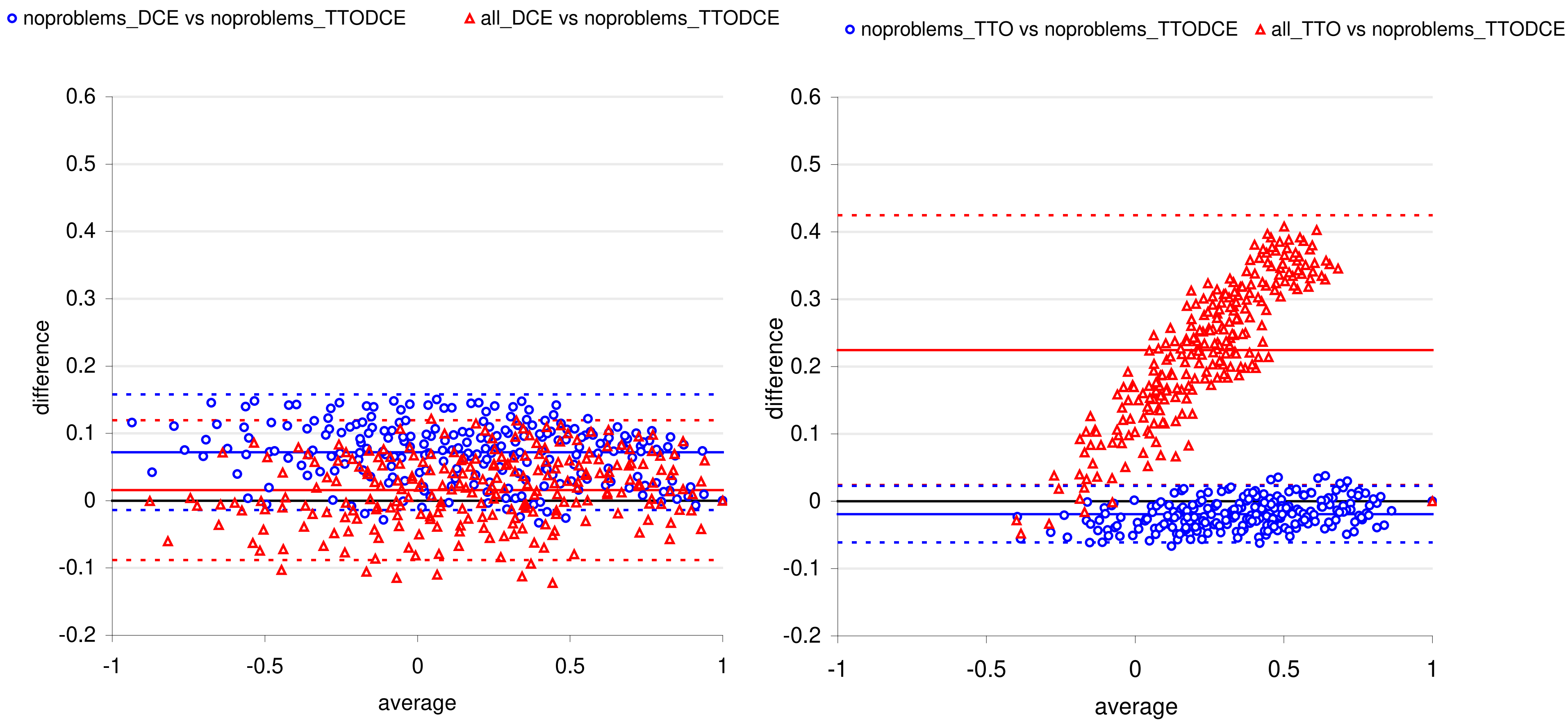


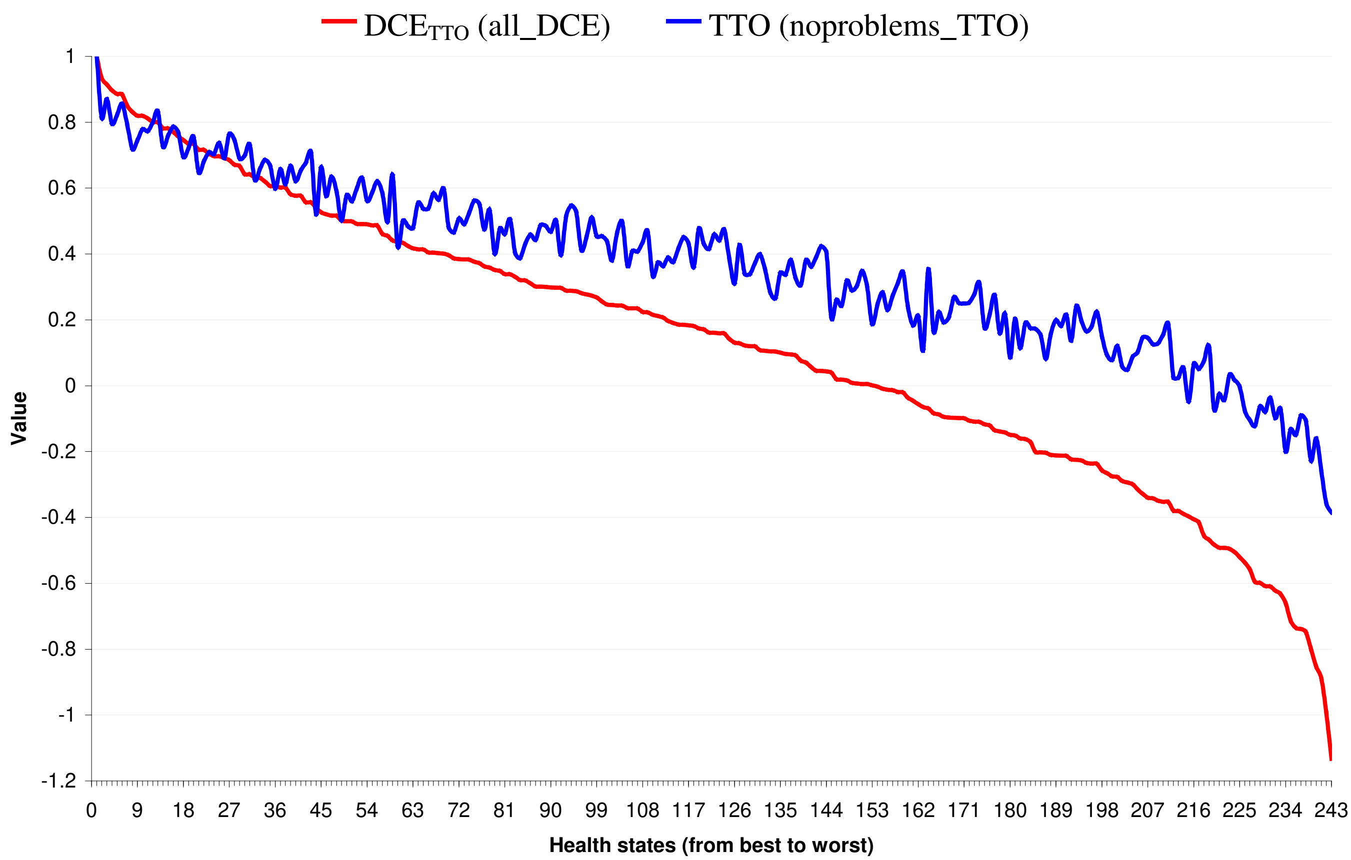
Figure 2. Bland Altman plots comparing means and differences in the estimated 243 EQ-5D values from the various TTO and DCE$_{TTO}$ samples

a) TTO b) DCE$_{TTO}$



Legend. Each figure shows the plot of the average versus the difference in each of the 243 values. The solid line indicates the bias (or MAD) and the dotted lines indicate the confidence intervals on the bias (1.96xSD of the difference)

Figure 3. A comparison between estimated values of the 243 EQ-5D health states for the TTO (noproblems_TTO) and DCE$_{TTO}$ (all_DCE)



Ordered by DCE$_{TTO}$ values

Appendix 1. All predicted values (TTO noproblems_TTO and DCE_{TTO} all_DCE)

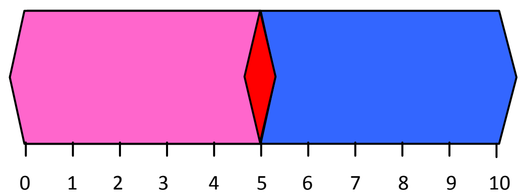
Health state	TTO	DCE_{TTO}	Health state	TTO	DCE_{TTO}	Health state	TTO	DCE_{TTO}
11111	1.000 (0.024)	1.000 (0.000)	32111	0.506 (0.033)	0.338 (0.033)	12233	0.109 (0.031)	-0.064 (0.035)
12111	0.815 (0.027)	0.935 (0.022)	12231	0.403 (0.031)	0.330 (0.033)	31322	0.354 (0.032)	-0.069 (0.032)
21111	0.870 (0.027)	0.915 (0.021)	12223	0.388 (0.031)	0.321 (0.037)	21233	0.165 (0.029)	-0.084 (0.032)
11211	0.796 (0.028)	0.897 (0.024)	12132	0.430 (0.031)	0.319 (0.031)	23231	0.223 (0.028)	-0.087 (0.033)
11112	0.823 (0.026)	0.885 (0.021)	21231	0.459 (0.028)	0.311 (0.032)	31213	0.193 (0.029)	-0.094 (0.034)
11121	0.856 (0.027)	0.885 (0.022)	21223	0.443 (0.030)	0.302 (0.033)	23223	0.208 (0.029)	-0.097 (0.032)
22111	0.792 (0.028)	0.850 (0.027)	31211	0.487 (0.030)	0.300 (0.027)	31131	0.269 (0.034)	-0.097 (0.039)
12211	0.718 (0.029)	0.831 (0.033)	21132	0.486 (0.029)	0.300 (0.030)	33211	0.252 (0.031)	-0.098 (0.034)
12112	0.745 (0.028)	0.820 (0.031)	23212	0.468 (0.029)	0.299 (0.033)	23132	0.250 (0.029)	-0.099 (0.030)
12121	0.778 (0.029)	0.819 (0.029)	23221	0.502 (0.029)	0.298 (0.034)	31123	0.253 (0.033)	-0.106 (0.041)
21211	0.773 (0.027)	0.812 (0.029)	12313	0.397 (0.030)	0.297 (0.032)	33112	0.279 (0.036)	-0.109 (0.037)
21112	0.801 (0.028)	0.800 (0.026)	31112	0.514 (0.035)	0.289 (0.031)	33121	0.312 (0.034)	-0.110 (0.039)
21121	0.834 (0.029)	0.800 (0.031)	31121	0.547 (0.034)	0.288 (0.035)	13232	0.176 (0.029)	-0.117 (0.033)
11212	0.727 (0.028)	0.782 (0.032)	23122	0.529 (0.031)	0.287 (0.032)	23313	0.217 (0.030)	-0.121 (0.033)
11221	0.760 (0.028)	0.781 (0.031)	11232	0.412 (0.028)	0.281 (0.032)	32322	0.276 (0.031)	-0.135 (0.033)
11122	0.787 (0.029)	0.770 (0.029)	21313	0.452 (0.031)	0.277 (0.029)	11333	0.160 (0.030)	-0.138 (0.033)
11311	0.769 (0.030)	0.757 (0.021)	23311	0.511 (0.032)	0.274 (0.032)	13331	0.219 (0.031)	-0.142 (0.032)
22211	0.695 (0.027)	0.746 (0.034)	13222	0.455 (0.031)	0.268 (0.035)	22233	0.086 (0.029)	-0.149 (0.032)
22112	0.722 (0.029)	0.735 (0.031)	11331	0.454 (0.032)	0.256 (0.028)	13323	0.203 (0.030)	-0.151 (0.031)
22121	0.756 (0.030)	0.734 (0.035)	11323	0.438 (0.031)	0.247 (0.029)	32213	0.115 (0.029)	-0.160 (0.036)
12212	0.648 (0.028)	0.717 (0.039)	22231	0.381 (0.028)	0.245 (0.034)	32131	0.191 (0.035)	-0.162 (0.040)
12221	0.682 (0.029)	0.716 (0.037)	13312	0.464 (0.033)	0.244 (0.032)	32123	0.175 (0.033)	-0.172 (0.042)
12122	0.709 (0.031)	0.705 (0.036)	13321	0.497 (0.033)	0.243 (0.029)	31231	0.172 (0.030)	-0.200 (0.035)
21212	0.704 (0.026)	0.697 (0.032)	22223	0.365 (0.029)	0.236 (0.035)	23232	0.154 (0.025)	-0.202 (0.031)
21221	0.737 (0.028)	0.696 (0.036)	32211	0.409 (0.030)	0.235 (0.031)	12333	0.082 (0.030)	-0.204 (0.034)
12311	0.691 (0.031)	0.691 (0.028)	22132	0.408 (0.030)	0.234 (0.031)	31223	0.157 (0.029)	-0.210 (0.036)
21122	0.764 (0.030)	0.685 (0.034)	32112	0.436 (0.035)	0.224 (0.035)	31132	0.199 (0.035)	-0.211 (0.039)
21311	0.746 (0.031)	0.672 (0.029)	32121	0.469 (0.034)	0.223 (0.037)	33212	0.182 (0.032)	-0.212 (0.034)
11222	0.690 (0.028)	0.667 (0.036)	12232	0.334 (0.029)	0.216 (0.036)	33221	0.215 (0.030)	-0.213 (0.035)
11312	0.699 (0.031)	0.642 (0.028)	22313	0.374 (0.030)	0.212 (0.030)	21333	0.137 (0.030)	-0.223 (0.034)
11321	0.733 (0.032)	0.641 (0.027)	13113	0.363 (0.029)	0.207 (0.030)	33122	0.242 (0.036)	-0.225 (0.039)
22212	0.626 (0.026)	0.632 (0.036)	21232	0.389 (0.026)	0.196 (0.032)	23331	0.196 (0.030)	-0.227 (0.035)
22221	0.659 (0.027)	0.631 (0.038)	12331	0.376 (0.033)	0.191 (0.030)	31313	0.166 (0.027)	-0.234 (0.036)
22122	0.686 (0.030)	0.620 (0.037)	31212	0.417 (0.031)	0.186 (0.029)	23323	0.180 (0.030)	-0.236 (0.032)
22311	0.668 (0.031)	0.606 (0.031)	31221	0.451 (0.030)	0.185 (0.032)	33311	0.225 (0.030)	-0.238 (0.037)
11113	0.599 (0.026)	0.605 (0.022)	23222	0.432 (0.027)	0.183 (0.035)	13332	0.149 (0.030)	-0.257 (0.032)
13111	0.658 (0.029)	0.602 (0.020)	12323	0.360 (0.031)	0.181 (0.032)	32231	0.094 (0.030)	-0.266 (0.036)
12222	0.612 (0.029)	0.601 (0.041)	31122	0.478 (0.036)	0.174 (0.036)	32223	0.078 (0.028)	-0.275 (0.037)
21222	0.667 (0.026)	0.582 (0.037)	21331	0.431 (0.032)	0.171 (0.033)	32132	0.121 (0.036)	-0.277 (0.039)
12312	0.621 (0.031)	0.577 (0.034)	21323	0.416 (0.032)	0.162 (0.032)	22333	0.059 (0.030)	-0.289 (0.032)
12321	0.654 (0.032)	0.576 (0.031)	31311	0.460 (0.030)	0.161 (0.030)	13133	0.049 (0.028)	-0.294 (0.035)
21312	0.677 (0.031)	0.557 (0.030)	23312	0.441 (0.031)	0.159 (0.032)	32313	0.088 (0.025)	-0.299 (0.037)
21321	0.710 (0.033)	0.556 (0.034)	23321	0.474 (0.032)	0.158 (0.033)	31232	0.103 (0.030)	-0.315 (0.034)
12113	0.521 (0.028)	0.540 (0.029)	11332	0.385 (0.031)	0.142 (0.030)	33222	0.146 (0.030)	-0.328 (0.034)
11322	0.663 (0.032)	0.527 (0.032)	22232	0.311 (0.026)	0.131 (0.033)	31331	0.145 (0.030)	-0.340 (0.039)
21113	0.576 (0.029)	0.520 (0.026)	13322	0.428 (0.033)	0.129 (0.032)	23332	0.126 (0.027)	-0.342 (0.032)
23111	0.635 (0.030)	0.517 (0.026)	23113	0.341 (0.030)	0.122 (0.030)	31323	0.129 (0.027)	-0.349 (0.037)
22222	0.589 (0.026)	0.516 (0.040)	32212	0.339 (0.030)	0.120 (0.033)	33312	0.155 (0.031)	-0.352 (0.036)
11213	0.502 (0.030)	0.502 (0.029)	32221	0.373 (0.029)	0.120 (0.035)	33321	0.188 (0.029)	-0.353 (0.036)
11131	0.578 (0.028)	0.499 (0.021)	32122	0.400 (0.036)	0.108 (0.038)	23133	0.026 (0.028)	-0.379 (0.035)
13211	0.561 (0.032)	0.498 (0.030)	22331	0.353 (0.032)	0.106 (0.033)	32232	0.025 (0.030)	-0.380 (0.035)
22312	0.599 (0.030)	0.492 (0.033)	11133	0.284 (0.029)	0.105 (0.030)	33113	0.054 (0.032)	-0.389 (0.045)
22321	0.632 (0.032)	0.491 (0.035)	13213	0.267 (0.032)	0.104 (0.034)	13233	-0.048 (0.029)	-0.397 (0.035)
11123	0.562 (0.029)	0.490 (0.028)	13131	0.343 (0.030)	0.101 (0.027)	32331	0.067 (0.029)	-0.405 (0.038)
13112	0.588 (0.031)	0.487 (0.027)	22323	0.338 (0.031)	0.096 (0.032)	32323	0.051 (0.026)	-0.415 (0.036)
13121	0.621 (0.030)	0.486 (0.025)	32311	0.382 (0.029)	0.095 (0.031)	31332	0.075 (0.030)	-0.455 (0.036)
12322	0.585 (0.032)	0.461 (0.036)	13123	0.327 (0.030)	0.092 (0.031)	33322	0.118 (0.030)	-0.468 (0.034)
22113	0.498 (0.030)	0.455 (0.029)	12332	0.306 (0.031)	0.076 (0.033)	23233	-0.071 (0.026)	-0.482 (0.033)
21322	0.640 (0.032)	0.442 (0.035)	31222	0.381 (0.031)	0.070 (0.033)	31133	-0.025 (0.032)	-0.491 (0.047)
12213	0.424 (0.030)	0.437 (0.035)	21332	0.362 (0.030)	0.057 (0.032)	33213	-0.042 (0.029)	-0.492 (0.040)
12131	0.500 (0.031)	0.434 (0.027)	31312	0.390 (0.031)	0.046 (0.030)	33131	0.034 (0.033)	-0.495 (0.045)
12123	0.484 (0.031)	0.425 (0.033)	31321	0.423 (0.031)	0.045 (0.033)	33123	0.018 (0.031)	-0.505 (0.045)
21213	0.479 (0.029)	0.417 (0.029)	23322	0.405 (0.031)	0.044 (0.032)	32332	-0.003 (0.029)	-0.520 (0.035)
21131	0.555 (0.029)	0.414 (0.029)	12133	0.206 (0.031)	0.039 (0.033)	13333	-0.075 (0.027)	-0.537 (0.036)
23211	0.538 (0.030)	0.413 (0.032)	21133	0.261 (0.030)	0.020 (0.033)	32133	-0.103 (0.033)	-0.557 (0.047)
21123	0.540 (0.031)	0.405 (0.032)	23213	0.244 (0.030)	0.019 (0.032)	31233	-0.122 (0.028)	-0.595 (0.041)
31111	0.584 (0.032)	0.404 (0.029)	23131	0.320 (0.029)	0.016 (0.032)	33231	-0.063 (0.028)	-0.599 (0.040)
23112	0.565 (0.031)	0.402 (0.028)	31113	0.290 (0.032)	0.009 (0.038)	33223	-0.079 (0.027)	-0.608 (0.039)
23121	0.598 (0.031)	0.401 (0.031)	23123	0.304 (0.031)	0.007 (0.033)	33132	-0.036 (0.034)	-0.610 (0.043)
11231	0.481 (0.029)	0.396 (0.028)	33111	0.348 (0.034)	0.005 (0.037)	23333	-0.098 (0.025)	-0.622 (0.036)
11223	0.466 (0.030)	0.387 (0.032)	32222	0.303 (0.030)	0.005 (0.035)	33313	-0.070 (0.024)	-0.632 (0.043)
11132	0.508 (0.029)	0.385 (0.026)	11233	0.187 (0.030)	0.001 (0.032)	32233	-0.200 (0.028)	-0.660 (0.040)
13212	0.491 (0.032)	0.384 (0.035)	13231	0.246 (0.031)	-0.002 (0.031)	33232	-0.133 (0.028)	-0.713 (0.037)
13221	0.524 (0.031)	0.383 (0.032)	22332	0.284 (0.029)	-0.009 (0.032)	31333	-0.149 (0.025)	-0.735 (0.044)
22322	0.562 (0.031)	0.376 (0.036)	13223	0.230 (0.031)	-0.012 (0.033)	33331	-0.090 (0.026)	-0.738 (0.043)
13122	0.551 (0.032)	0.372 (0.030)	13132	0.273 (0.030)	-0.014 (0.029)	33323	-0.106 (0.023)	-0.748 (0.041)
11313	0.475 (0.030)	0.362 (0.027)	32312	0.312 (0.030)	-0.019 (0.032)	32333	-0.227 (0.024)	-0.800 (0.042)
13311	0.534 (0.033)	0.359 (0.028)	32321	0.345 (0.030)	-0.020 (0.033)	33332	-0.160 (0.026)	-0.853 (0.040)
22213	0.401 (0.029)	0.352 (0.032)	13313	0.240 (0.030)	-0.036 (0.033)	33133	-0.261 (0.029)	-0.890 (0.051)
22131	0.477 (0.031)	0.349 (0.031)	22133	0.183 (0.031)	-0.046 (0.033)	33233	-0.357 (0.024)	-0.993 (0.045)
22123	0.462 (0.032)	0.340 (0.034)	32113	0.212 (0.032)	-0.056 (0.040)	33333	-0.384 (0.017)	-1.133 (0.048)

Appendix 2. Task examples

TTO Better than dead task example

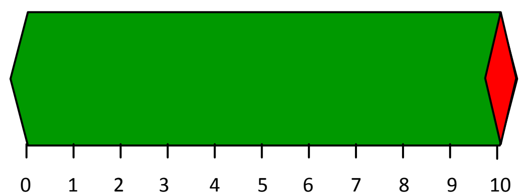
Now you would either live in Life A for 5 years and then die, or you would live in Life B for 10 years and then die. Would you prefer Life A or Life B, or are they the same?

- ☐ No problems in walking about
- ☐ No problems with self-care
- ☐ No problems with performing usual activities
- ☐ No pain or discomfort
- ☐ Not anxious r depressed



LIFE A

- ☐ Some problems in walking about
- ☐ No problems with self-care
- ☐ No problems with performing usual activities
- ☐ Moderate pain or discomfort
- ☐ Extremely anxious or depressed



LIFE B

Choose one	LIFE A ☐	LIFE B ☐	THE SAME ☐

DCE$_{TTO}$ task example

Now you would either live in Life A for the described number of years and then die or live in Life B for the described number of years and then die. Would you prefer Life A or Life B?

	LIFE A	**LIFE B**
Anxiety/depression	Extremely anxious or depressed	Not anxious or depressed
Pain/discomfort	Moderate pain or discomfort	Extreme pain or discomfort
Mobility	Confined to bed	No problems in walking about
Usual Activities	Some problems performing usual activities	Some problems performing usual activities
Self-care	Unable to wash or dress self	No problems with self-care
Duration of life	Live for 4 years	Live for 10 years
Choose one	☐	☐