A Comparison of a Discrete Choice Experiment to the Time-Trade-Off to value health states for QALYs.

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

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

1. Introduction

Understanding how people value health benefits has become a major field of health economics research. Information from such studies has been advocated to inform resource allocation decisions in health care. Economists typically measure the value of commodities by examining the prices of goods and services bought and sold in the marketplace. Given the existence of market failure in health care, it is difficult to use market data to inform the value of health benefits. Instead, benefits have generally been valued relating to an individual's strength of preference where the stronger an individual's preference for a particular outcome, the greater the value of the benefit. Various methods have been developed for valuing states of health, known as health utilities, which reflect individual's preferences (1,2). The Quality Adjusted Life Year (QALY) is a widely adopted metric which enables health-care decision-makers to weigh the implementation or purchase of health-care technologies, including diagnostics, devices, and medications. The QALY is derived by assigning a health utility to each state of health based on its value, multiplied by how long the state lasts.

The increasing use of QALYs to inform decision making has led to greater scrutiny of the underpinning methods and underlying theory – in particular the methods for valuing health states. Health utilities obtained from these methods must meet several conditions to be applicable for use in QALY calculations (3). Firstly, the values should be based on preferences as opposed to psychometric methods where numerical assessments are used to reflect individuals' health status. Secondly, the values must be measured on a cardinal scale since incremental gains in QALYs are compared across programs or interventions. Finally, the values must be anchored on measures of 1 (full health) and 0 (death) so that the absence of a year is equivalent to a QALY of zero. The most widely used methods for health state valuation include the Time Trade Off (TTO), Standard Gamble (SG) and the Visual Analogue Scale (VAS). Each method has been utilized in large stated preference surveys where members of the general population or patients are asked to value health states that they may or may not have experienced.

A common feature of all three conventional elicitation methods is that a respondent must describe their preference or rating on an absolute cardinal scale. For the TTO and SG this is done iteratively by responding to a set of binary choices until reaching a point of indifference which indicates the degree that one health state is preferred to another. For the VAS this is simply a rating on a scale. The psychological and cognitive aspect for eliciting preferences in such a way has been criticized for, amongst other aspects, framing effects and loss aversion (4,5). Recent evidence from the psychology literature provides important insights into the heuristics a respondent uses in preference construction. It suggests the human brain does not have an internal scale and so when people try to represent a magnitude, they can do so only on the basis of whether it is larger or smaller than other magnitudes retrieved from memory or observed in the environment (various experiments have been reported including asking people to rate how light it is based on the anchors of pure light to dark (6)). Such evidence gives credence to the focus in recent years that proposes the use of ordinal data collection techniques as an alternative to the conventional cardinal elicitation techniques (7-12). Such techniques require respondents to rank responses such as stating that A is preferred to B (without reference to any numerical trade-off). Proponents argue that given the strong methodological and psychological foundation for ordinal techniques, elicitation using such methods will lead to more consistent and reliable responses.

One ordinal elicitation method that is increasingly being used in health economics is the discrete choice experiment (DCE). Such experiments involve the construction of a set of alternatives based upon a limited number of important attributes and then obtaining preferences over those alternatives from respondents. The exercise can be repeated with alternatives comprising different levels of the attributes in order to infer the relative weight attached to each level of each attribute. It is able to generate information to indicate the relative importance of levels between and within attributes (13). While the appeal to using DCE’s as an alternative to conventional techniques does appear to have some methodological and theoretical basis, it has been cautioned to consider equally the limitations that DCE’s bring over conventional elicitation techniques (14,15).

Little research has been conducted in the comparison between DCE’s and other preference elicitation techniques. Consequently, whether respondents give more consistent and reliable assessment of health state values has not been formally assessed. While an ordinal task may be less cognitively demanding, facing two or more varying health states instead of one in each task (e.g. full health does not change between scenarios) can still be challenging, particularly when there are many attributes. Some evidence suggests respondents faced with such complex problems tend to adopt simple decision heuristic strategies (16,17). Another issue is whether ordinal methods demand more data in order to obtain a similar variance in parameter estimates in comparison to conventional techniques. If more data is required, improved cognition to a task might be offset against the cognitive burden by respondents having to spend longer on surveys, or by requiring more resources to recruit respondents to participate. A final issue surrounds the methodological challenges that exist for anchoring results of DCE's on full death and death. Since ordinal methods only provide information on the relative preference of one health state to another, it is necessary to scale estimates so that these preferences are relative death and full health.

This paper explores the use of a DCE as an alternative to the conventional TTO technique to value health states for one of the most widely used health classification systems, the EQ-5D. In the first section the literature on previous DCE exercises to value health states is examined to guide the design of a DCE for deriving health utilities. In the second section the valuation study which elicited stated preferences using both a DCE and a TTO for a health state classification system is described. Estimated health utilities for each health state within the classification are then estimated using different statistical models for each of the two methods. The results of each method are then compared in section 3. Finally the implications of these results for future elicitation of health state values are discussed.

2. Using DCE's to value health profiles

An update of previous reviews of the literature found that the use of DCE's in the health economic literature is steadily rising (18, 19). While these papers cover a wide range of topics, a vast majority have sought to assess the benefits of specific health care services. Such studies typically focus on the health gains from different interventions, where the various benefits and sometimes side-effects of health are used as attributes. Sometimes, non-health outcomes and 'process' factors have been included. By including a financial cost as an attribute, the implied willingness to pay for different levels of attributes is calculated. The search reveals that only 12 studies so far have looked to use DCE's for the objective of only valuing health profiles. Only 5 of these studies had a similar objective to this study – to use ordinal data derived from DCE's to estimate cardinal health utilities anchored on death and full health scale.

Three of the studies used a standard pairwise choice design consisting of attributes and levels from health status instruments, and then used external anchors to rescale values to the 0-1 scale derived from the results of the statistical analysis.(7,20,21).The studies by Ryan and Burr used a similar assumption - that the best possible health state is equivalent to full health and the worst health state is equivalent to death. The study by Ratcliffe overcomes this arbitrary scaling by using elicitations from the same individuals for the values of the best and worst health states using a TTO. While this scaling improves the theoretical basis for the values, the reliance on using TTO elicitations contradicts the motivation of using DCE's instead of conventional techniques to value health states.

Two studies used an alternative method where death is included as an alternative in each choice set (8; 22). Brazier et al used a pairwise DCE approach where a number of health states are paired with dead. Flynn et al used Best-Worst Scaling and asked respondents an additional question as to whether they thought it was a state worth living. The Flynn approach enables the value of the worst health state (and therefore all other states) is estimated relative to death. Brazier et al used the same approach as has been applied to rank data and estimated coefficients for a 'dead dummy' alongside the various attributes and then used this to transform the coefficients onto the full health dead scale. While an attractive solution to the issue of rescaling, initial findings by Flynn suggests that the lack of individuals choosing death in any choice set undermines the some assumption of utility theory used in the statistical analysis of such data. Whether death is chosen depends on the alternative health state(s) and the study by Brazier et al which includes arguably more severe health states was able to estimate use the approach.

A third approach has been proposed in which survival is included as an attribute (rather than the state of being dead) (8; 23; 24). This results in a DCE that reflects more closely the TTO by asking for choices between health states that include an additional attribute with levels depicting the length of life that would be lived in the particular state has been suggested as another alternative approach - potentially combining the cognitive advantage of ordinal choices with assumptions already studied from existing elicitation techniques. Given the multiplicative nature of health status and life years in the QALY function, the main drawback to this is that standard additive designs would likely not suffice, and so the use of an experimental design that can estimate interaction effects would be required. It is this approach that is studied in this paper.

3. Methods

3.1 Valuation protocol

The EQ-5D health state classification system was used as the instrument to populate with preference weights (25). This was chosen because it has been the focus of many previous valuation surveys, and also because it has one of the smallest number of health states. The EQ-5D descriptive system consists of 5 dimensions: Mobility (MO), Self-Care (SC), Usual Activities (UA), Pain/Discomfort (PD), and Anxiety/Depression (AD). Each dimension has 3 levels, reflecting "no health problems," "moderate health problems," and "extreme health problems." A dimension for which there are no problems is said to be at level 1, whereas a dimension for which there are extreme problems is said to be at level 3. Each unique health state described by the instrument has an associated 5-digit descriptor ranging from 11111 for perfect health to 33333 for the worst possible state (26).

3.1.1. Time Trade Off

The TTO starts with the premise that health is an important argument in an individual's utility function. The welfare change associated with a change in health status is determined by valuing the amount of life expectancy an individual is prepared to sacrifice that leaves overall utility (defined as Π) unchanged (27). The approach involves presenting an individual with a paired comparison. The first alternative involves living for period *T* in a specified but in a less than full health state (state *h*). The second alternative involves living in full health (state *FH*) for a period of time *t* where $t<T$. The health utility *U* for a given health state *h* is calculated by varying *t* until the overall preferences between the two alternatives are equal:

$$\Pi[U(FH), t] = \Pi[U(h), T] \tag{1}$$

Different protocols exist for conducting TTO studies (28). It was chosen to follow broadly to follow the TTO elicitation section of the MVH protocol so to replicate previous elicitation surveys.(29; 30; 26) In this survey respondents were given health states made up of the 5 EQ-5D attributes and asked to assume that the duration of each health state was 10 years ($T=10$), followed by immediate death. A choice is then presented: to live in the given health state *h* or immediate death. If the living health state is chosen, tradeoffs are made using a visual board to determine the number of years in full health (11111) that is equivalent to 10 years in the state presented. If the respondent prefers death over the health state presented, the survey proceeds by asking if the respondent prefers 5 years in the given state followed by 5 years in full health, or again death. The time *t* in the health state is then varied by 1 year and then 6 monthly intervals, based on responses, until the point of indifference is found. In the first task, respondents are asked whether they would prefer living in full health for 10 years or the given health state for 10 years. The results of this question are used as one of the consistency checks.

TTO data produces values directly meeting the criteria for health utility values. For states better than death, the individual's health state value is defined as the point of indifference *t* divided by 10, while for states worse than death, following Shaw and colleagues, the individual's health state value was calculated by dividing the point of indifference *t* by a constant representing the greatest negative value in the range.(30)

3.1.2. Discrete Choice Experiment

The economic theory behind DCE's is based on random utility maximization (RUM) framework.(31; 32) If we let Π_{ij} be the overall utility assigned by individual *i* to health state *j*, *j=1,…,C* , in a choice set with *C* options, then option *m* is chosen if an only if $\Pi_{im} > V_{ij} \; \forall j \neq m$.

However, in a behavioral experiment such as a DCE, the researcher does not see the individuals overall utility, but only the options offered and the choices made. If we define the observable systematic component of the utility as V_{ij}, we can assume that

$$\Pi_{ij} = V_{ij} + \varepsilon_{ij} \tag{2}$$

where ε_{ij} includes all the things that have affected overall utility but cannot be observed.

Then, we can say we know that:

$$\begin{aligned} Pr(\text{alternative } m \text{ is chosen by subject } i) &= Pr(\Pi_{im} > \Pi_{ij} \; \forall j \neq m) \\ &= Pr(V_{im} + \varepsilon_{im} > \Pi_{ij} + \varepsilon_{im} \; \forall j \neq m) \\ &= Pr(V_{im} - V_{ij} > \varepsilon_{im} - \varepsilon_{ij} \; \forall j \neq m) \end{aligned} \tag{3}$$

If the error terms are assumed to be independent and identically distributed with an extreme value distribution, the probability of the *j*th scenario being chosen from the complete choice set (m=1 to n) can be characterized using the conditional logit model (sometimes referred to as the multinomial logit (MNL) model) as the following:

$$Pr(U_{ij}) = \frac{e^{(V_{ij})}}{\sum_{m=1}^{M} e^{(V_{im})}} \tag{4}$$

The protocol adopted for the DCE uses the 5 attributes for the EQ-5D, but adds a 6 attribute 'life years' so to observe preferences over how much life expectancy individuals would sacrifice for living in different states of health. Consequently in this exercise, V_{ij} can be seen as a function of health utility *U* and time *t.* Reflecting the TTO procedure, the life years attribute ranged from 1 to 10 years, with levels at 4 and 6 years[1]. Individuals were shown two health profiles in each choice set, each describing living in a particular EQ-5D health state for a particular number of years. Respondents were asked to simply choose which health profile they would prefer (e.g. a forced choice).[2]

3.2 Experimental design

---

[1] The use of zero years was excluded after pilot work based on qualitative responses of respondents and due to the assumptions required for modeling the data

[2] In pilot testing, where a third choice of 'the same' was added, it was frequently chosen and when asked why, it was determined that often respondents did not want to think through the choice, and so this was not included in the main study

The EQ-5D descriptive system defines 243 (3^5) health states. Based on results of a pilot study, it was determined that respondents would be able to value 5 health states in the TTO exercise, and 8 DCE choice tasks (but 2 of these would be consistency checks). For both the TTO and DCE, it was considered impractical to value all combinations of attributes and levels, and so it was decided to include a subset of states based on published orthogonal arrays in order to obtain non confounded estimates of effects. The array for the TTO was smaller to that of the DCE, as it only required the estimate of main effects while the DCE required the estimation of at least the interaction between each of the 5 EQ-5D attributes and the 'life years' attribute. While smaller orthogonal arrays exist, an array of 36 states was chosen so to allow more comparisons with states from the DCE. The arrays were chosen so that they included only realistic health states, and that they overlapped as far as possible. Additional states were chosen so that all 17 recommended by Lamers et al were directly valued (33). This would enable, for the TTO, results to be compared with values from many previous valuation studies. A total number of 48 health states were consequently valued.

For the DCE experimental design theory was consulted (34; 35). A number of near orthogonal designs were generated. Choice sets were the one with the fewest dominated choice sets, highest overlap with choice sets where both states were valued in the TTO, and highest D-efficiency (asymptotic variance–covariance matrix of the parameter estimates) was selected and tested using the macros from Kuhfeld.(36).[3] In the end, the design with 144 choice sets was selected.

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 level balance was achieved between sets. The DCE was blocked in a similar way into 24 sets and each matched with one of the TTO sets (e.g. 2 DCE sets to each TTO set) such that there was overlap where possible between states in the two sets.

3.3 Valuation survey

The aim was to elicit values for health states from a representative sample of 1500 adult members of the general population. By recruiting 1500 respondents, this would give approximately 500 estimates of each health state from the TTO – similar to existing surveys, and over 50 estimates of each choice set in the DCE. Given the resources for the project, it was determined that interviews were not feasible and that the web was the best method to administer the survey. A market research company was hired to recruit individuals greater than 18 years of age who were residents of British Columbia, Canada. Initial

[3] The D-efficiency of the 'optimal' was not able to be calculated and so it is likely a more efficient design exists

contact was made via email. Individuals choosing to participate in the study were referred to a password-protected website that contained the questionnaire. The market research company offered incentives to participants who completed the questionnaire. Quota sampling was used to obtain a representative sample of the Canadian population with regard to age and gender. A small pilot study (n=35) 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.

Prior to the elicitation of health state values using the TTO and DCE methods, the respondents were asked to describe their own health using the EQ-5D. In the second stage of the interview, the TTO or DCE task was randomly chosen. In each, a description of the task was given, and a demonstration video provided. In both surveys, the order of attributes was randomized (apart from 'life years' in the DCE) and the order of tasks randomized to test whether this influenced choices. To test consistency in the DCE, two questions were included where one of the two options in the choice set was regarded as a dominant option (i.e. all attribute levels were regarded as more desirable). At the end of the tasks, respondents were asked the difficulty with understanding and answering the previous tasks. In the third stage of the interview, the task not used in the second stage was used. The process was identical. In the fourth stage of the interview, the comparative difficulty with understanding and answering the DCE and TTO questions was assessed. We also directly asked respondents which technique they preferred. An opportunity to explain responses was given with an open ended question. Finally, interviewees were asked whether they would be willing to participate in a further internet survey.

To satisfy ethical requirements, in each of the second or third stages of the survey, respondents could opt out. If this happened, they were asked why (felt uncomfortable, lack of understanding, difficultly, boredom), and proceeded onto the next stage of the survey (which they could opt out of again). Socio-demographic questions (e.g. age, sex, employment status, household characteristics) were not asked as these are known to the market research company and were later linked with results of the survey.

3.4 Derivation of health state values

3.4.1 Time Trade Off analysis

The health utility values U_{ij} for each health state are directly obtained from each respondent by calculating t/T from each TTO scenario. Assuming a linear model (which is commonly done), the general model for estimating all health states of a descriptive system from conventional valuation data such as the TTO is (37):

$$U_{ij} = \left(\frac{t}{T}\right)_{ij} = \alpha + \boldsymbol{\beta'}\boldsymbol{x}_{ij} + \boldsymbol{\theta'}\boldsymbol{r}_{ij} + \boldsymbol{\delta'}\boldsymbol{z}_i + \varepsilon_{ij} \tag{5}$$

where i = 1, 2,...,n represents individuals and j = 1, 2,...,m represents the different health states shown to each respondent. The dependent variable, U_{ij} is the health utility obtained from the TTO for each respondent i for health state j. $\boldsymbol{x}_{ij}$ is a vector of 10 binary dummy explanatory variables where:

$$x_{ij}^{\delta\lambda} = \begin{cases} 1 & \text{where } \delta=1,2,\ldots,5 \text{ is the attribute and } \lambda=2,3 \text{ indicates the worse 2 levels of each health state } x \text{ of the } i^{th} \text{ individual, } j^{th} \text{ scenario} \\ 0 & \text{in all other situations} \end{cases}$$

Hence $\boldsymbol{\beta'}$ is vector of 10 variables ($\boldsymbol{\beta}^{12}, \boldsymbol{\beta}^{13}, \ldots, \boldsymbol{\beta}^{53}$). $\boldsymbol{r}_{ij}$ is a vector which accounts for possible interactions between the levels of different attributes and $\boldsymbol{z}_i$ is a vector of personal characteristics that may also affect the value an individual gives to a health state, for example, age, sex and education. Finally ε_{ij} is an error term whose autocorrelation structure and distributional properties depend on the assumptions underlying the particular model used. An important assumption to this modeling strategy is that the intercept α is restricted to a value of 1, so that in a linear model, full health (level 1 for each dimension, 11111), is the upper anchor of the TTO task.

A one-way error components random effects (RE) model was used to analyse the data. Following previous analysis of the EQ-5D, the dependent variable was modeled as 1 minus the health state (33).

3.4.2 Discrete Choice analysis

In our analysis each respondent is shown k scenarios each with 2 alternatives. We specify that each individual's latent utility function Π_{ijk} (detailed in equation 2) is made up of the main effects associated with each attribute and level associated with each health state $\boldsymbol{x}_{ijk}$, time t_{ijk}, possible interactions between health state levels $\boldsymbol{r}_{ij}$ and personal characteristics $\boldsymbol{z}_i$, as well as the interaction between each health state attribute level and time:

$$\Pi_{ijk} = \alpha + \boldsymbol{\beta_1}'\boldsymbol{x}_{ijk} + \beta_2 t_{ijk} + \boldsymbol{\beta_3}'\boldsymbol{x}_{ijk} \cdot t_{ijk} + \boldsymbol{\theta'}\boldsymbol{r}_{ij} + \boldsymbol{\delta'}\boldsymbol{z}_i + \varepsilon_{ijk} \tag{6}$$

where i=1,...,n; j=1,2; and k=1,...,K;

Therefore Π_{ijk} represents the utility that individual i derives from alternative health state j in scenario k. $\boldsymbol{x}_{ijk}$ again represents a vector of 10 binary dummy explanatory variables describing each potential EQ-5D health state. The t_{ijk} term specifies the level shown in the attribute life years but in the model is assumed to be a continuous.

Using a conditional logit model the predicted probabilities and coefficients for choosing alternative *j* in any given scenario can be estimated as:

$$\hat{p}(j|scenario) = \frac{e^{\alpha+\hat{\boldsymbol{\beta}}_1'\boldsymbol{x_j}+\hat{\beta}_2 t_j+\hat{\boldsymbol{\beta}}_3'\boldsymbol{x_j}\cdot t_j+\hat{\boldsymbol{\theta}}'\boldsymbol{r_j}+\hat{\boldsymbol{\delta}}'\boldsymbol{z_j}}}{e^{\alpha+\hat{\boldsymbol{\beta}}_1'\boldsymbol{x_1}+\hat{\beta}_2 t_1+\hat{\boldsymbol{\beta}}_3'\boldsymbol{x_1}\cdot t_1+\hat{\boldsymbol{\theta}}'\boldsymbol{r_1}+\hat{\boldsymbol{\delta}}'\boldsymbol{z_1}}+e^{\alpha+\hat{\boldsymbol{\beta}}_1'\boldsymbol{x_2}+\hat{\beta}_2 t_2+\hat{\boldsymbol{\beta}}_3'\boldsymbol{x_2}\cdot t_2+\hat{\boldsymbol{\theta}}'\boldsymbol{r_2}+\hat{\boldsymbol{\delta}}'\boldsymbol{z_2}}} \tag{7}$$

However, the estimates coefficients are not anchored on full health and death but rather relative to each other. To scale values to the full health death scale the following assumption is employed.

Using the estimated coefficients, equation 7 can be populated for a hypothetical scenario where in a pariwise choice set, the first alternative is living in full health for time *t*, and the second alternative is living in a given health state for time *T*. The objective is to reflect the point of indifference between these two alternatives as is directly measured in the TTO (equation 1), and so it is assumed the probability of choosing the first alternative is equal to the second. Since when ***x*** represents full health it is simply a vector of zero's[4], using the calculated coefficients this can be expressed as follows:

$$\frac{e^{\alpha+\hat{\beta}_2 t+\hat{\boldsymbol{\theta}}'\boldsymbol{r}+\hat{\boldsymbol{\delta}}'\boldsymbol{z}}}{e^{\alpha+\hat{\beta}_2 t+\hat{\boldsymbol{\theta}}'\boldsymbol{r}+\hat{\boldsymbol{\delta}}'\boldsymbol{z}}+e^{\alpha+\hat{\boldsymbol{\beta}}_1'\boldsymbol{x}+\hat{\beta}_2 T+\hat{\boldsymbol{\beta}}_3'\boldsymbol{x}\cdot T+\hat{\boldsymbol{\theta}}'\boldsymbol{r}+\hat{\boldsymbol{\delta}}'\boldsymbol{z}}} = \frac{e^{\alpha+\hat{\boldsymbol{\beta}}_1'\boldsymbol{x}+\hat{\beta}_2 T+\hat{\boldsymbol{\beta}}_3'\boldsymbol{x}\cdot T+\hat{\boldsymbol{\theta}}'\boldsymbol{r}+\hat{\boldsymbol{\delta}}'\boldsymbol{z}}}{e^{\alpha+\hat{\beta}_2 t+\hat{\boldsymbol{\theta}}'\boldsymbol{r}+\hat{\boldsymbol{\delta}}'\boldsymbol{z}}+e^{\alpha+\hat{\boldsymbol{\beta}}_1'\boldsymbol{x}+\hat{\beta}_2 T+\hat{\boldsymbol{\beta}}_3' x\cdot T+\hat{\boldsymbol{\theta}}'\boldsymbol{r}+\hat{\boldsymbol{\delta}}'\boldsymbol{z}}} \tag{8}$$

Assuming *T* is equal to 10, as it is in the TTO, only *t* is unknown on the left hand term and the given health state ***x*** is the only unknown on the right hand term. Consequently, by varying *t* until the two expressions are equal, we can solve *t/T* for each of the individual 243 states of *x*:

$$\frac{t}{T} = 1 + \left(\frac{\hat{\boldsymbol{\beta}}_1}{\hat{\beta}_2 T} + \frac{\hat{\boldsymbol{\beta}}_3}{\hat{\beta}_2}\right)' \boldsymbol{x} \tag{9}$$

3.6 Selection of valuation sample

It is common in surveys to receive a number of inconsistent or illogical responses from respondents. Such responses can be derived for many reasons – from respondents not understanding the task to situations where respondents simply complete the survey in order to gain the incentive. This latter scenario is of particular concern in a web survey where anonymity is increased in comparison to face to face interviews. It is difficult to know from response data which situation is occurring and so a number of reliability and consistency checks are used.

---

[4] Alternatively effects coding could be applied to the data and a value for $\boldsymbol{\beta}_1\boldsymbol{x}$ (***x***=full health) and $\boldsymbol{\beta}_3\boldsymbol{x}$.T (***x***=full health) could be applied, but it can be proven this ultimately gives the exact same result for the estimates of each health state.

Potential inconsistencies for the TTO were defined if the preliminary rationality question was answered incorrectly, if all the states were given the same value, if all states were valued worse than 'immediate death' or if there were 3 or more logical inconsistencies between states valued by the same respondent (e.g. for two states where the first has levels equal or worse for each of the attribute levels but the value of the first state is higher than the second) (38). Respondents were deemed inconsistent to the DCE if they answered one of the rationality checks incorrectly, if they chose all the health states (including dominated choices) on the left or the right, or if they answered within a time where it was deemed too quick for the respondent to read the task.

There is debate about whether too include or exclude inconsistent responses in the final valuation data. Given an objective of the study was to compare the values from each method, it was decided to initially exclude these responses from the valuation set.

The different valuation methods were compared with respect to; theoretical inconsistencies within levels of coefficients modeled for each attribute, the self-reported responses to questions regarding the difficulty of each task, the time taken to answer each task, and the actual values produced for each of the 243 health states.

4. Results

To date, only 691 respondents have participated in the study, and the preliminary results are shown below.

4.1 Sample size

A sample of 1786 member's of the market research panel was initially invited by email to participate in the survey. Of these 691 (39%) consented to begin the survey. In total, 554 (80% of those that consented) completed the survey, while 571 (83%) completed the TTO exercise and 588 (85%) completed the DCE exercise (7% during the first valuation exercise, 6% in the second). 38 respondents (6%) dropped out before beginning the first exercise and 14 (2%) dropped out during the last evaluation questions.

4.2 Data problems

Only 560 respondents completed both the DCE and TTO exercises. The various data problems within each valuation exercise are shown in Table 1. In total, 375 respondents (66% of the respondents completing the TTO) were defined to have no problems in their TTO valuations. 551 respondents (94% of respondents completing the DCE) were defined to have no problems in their DCE responses. It should be noted that since the criteria for problems differs between techniques, the numbers of data problems should not be directly compared (e.g. the TTO might have more stringent criteria). In this preliminary valuation analysis, it was decided to only include respondents with no problems who have completed both exercises (n=349, 62% of respondents that answered both exercises).

4.3 Characteristics

The characteristics of respondents are shown in Table 2. The valuation sample had a mean age of 54, some 5 years younger than the non-completes and 3 years younger than respondents with data problems. Levels of education, marital status and gender were not different between groups. The valuation sample had slightly higher income to both respondents that did not complete both tasks or those with data problems. The valuation sample is broadly representative of the population of Canada in terms of gender, age and education. EQ-5D responses in the valuation sample are very similar to a previous study EQ-5D (no problems in MO 78% vs 78%, SC 97% vs 96%, UA 80% vs 81%, PD 47% vs 56% and AD 72% vs 71%) (39).

When variables for characteristics of respondents were added in the regression models, neither were the variables statistically significant not did the coefficients of the variables describing the health states was not affected. These variables were consequently removed from the model.

4.4 DCE results

Table 3 shows the results from the conditional logit model. The life years attribute is the only main effect with a statistically significant coefficient. The coefficients of each of the level 3 interactions with life years were all significant with the exception of the usual activities coefficient. Using the assumptions described for equation 9, estimates for the 243 health states were derived using estimated coefficients. Resulting values ranged from 1.025 (state 11211) to -0.983 (state 33333). The corresponding coefficients from the DCE can be estimated (without a constant term) and are shown in Table 4. This shows that all attributes are logically consistent with the exception of level 2 of the usual activities attribute which is negative (suggesting it is better than level 1). Further analysis of results using a mixed logit is ongoing.

4.5 TTO results

The mean and standard errors of the 48 health states are presented in the appendix A. Mean TTO health state values ranged from -0.259 to 0.843 based on the Shaw et al transformation of the values for states worse than dead. Most health states had between 23 to 37 observations with the worst health state

(33333) having 349 (i.e. all the respondents). The results for the random effects model are shown in Table 4. All the coefficients were positive, and the coefficients increased in absolute terms in each attribute with the increasing level, which is as expected given each level represents progressively worse problems on each attribute compared with the baseline. The coefficients of the two levels depicting problems in usual activities were very close however. All of the attribute level coefficients were statistically significant with the exception of the level 2 coefficients of mobility and pain/depression ($p<0.05$), along with the constant term.

4.6 Comparison of estimated values

The estimated values for the 243 health states for the EQ-5D based on the TTO and DCE are shown in Figure 1 and in appendix B. There is divergence particularly in worse health states. The largest absolute difference is in health state 33333 where the TTO predicts a value of -0.301 while the DCE predicts a value of -0.983.

4.7 Comparison of responses

The mean time taken to complete the survey was just over 21 minutes (IQR 14-24). As expected respondents took more time in completing the first question of each exercise (over 2 and half 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)- see Table 5. In total, on average respondents took nearly 9 minutes to complete the 5 TTO valuations and under 4 minutes to complete the 8 DCE 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 over a half of respondents finding each exercise fairly or very difficult. However, respondents in the data problems group found the TTO more difficult to understand (but not to answer) than the valuation sample.

5. Discussion

While the study is still ongoing limiting the conclusions that can be made thus far, the results to date do demonstrate the feasibility of using a DCE to obtain health utilities for a widely used classification system. The values obtained from the DCE appear reasonable. The one attribute not theoretically consistent, causing values to be above 1, might change with the increased sample size. Given the hypothesis that DCE's might provide a simpler cognitive task to the TTO, there is little reason to expect models to generate the same values as those from the TTO. Consequently, the results of the paper focus on the consistency of results, the assumptions required to generate values, and the cognitive complexity and burden of the survey.

A possible advantage of this DCE variant over the TTO (and SG) is in the way it handles states worse than dead. In the TTO, the valuation task for states worse than dead is different to that for states better than dead. It results in values with very high negative values - that can in principle be negative infinity. For this reason analysts have typically applied arbitrary transformations (e.g. Shaw et al (30)). This DCE variant is able to cover all states better than dead or worse than dead in one task and does not require an arbitrary transformation to avoid extreme values, though the linear assumption of time requires further scrutiny.

A preliminary result was that respondents did not find the DCE simpler to understand or answer than the TTO. While the cognitive complexity of an ordinal choice might be simpler than picking a value on an ordinal scale, the DCE consisted of 2 health states which change with each task, each with varying life years whereas the TTO has just one health state that changes between tasks. Respondents completed the DCE quicker to the TTO. This might be advantageous – reducing respondent burden and survey costs, but of course less time might indicate less care, or less engagement.

There are a number of issues worthy of further consideration. First, the functional form of each analysis might be problematic. For the DCE, the main effect for ‘life years’ is significant suggesting that respondents had a preference for states with longer duration irrespective of the state itself – violating a fundamental condition for the QALY. It is not clear whether these main effects should be excluded from the model – ignoring the problem but leading to a problematic error term, or if it should remain included. Another issue for both the functional forms of the DCE and TTO is the interactions between attributes. The experimental designs of each do not allow these interactions to be estimated, but previous work including an N3 or D3 term might be useful to explore.

Second, the experimental design of the DCE was complicated given the need to estimate interactions between the life years attribute and the EQ-5D attributes. Since not all EQ-5D health states are logically possible, orthogonal designs are not possible, and so near orthogonal designs were generated which minimized the correlation between effects. There has been substantial progress made in the experimental design theory for DCE’s in the past few years, but so called ‘optimal’ designs were not able to be generated for this problem. Our design also considered other objectives that would not necessarily be considered if a de novo valuation study was being conducted. Consequently, if future studies using this DCE variant are being considered, the work on experimental design theory by Street and Burgess should be further studied (34).

Third, we excluded a number of respondent's values from our valuation sample since their responses seemed values seemed inconsistent or illogical. There is a substantial literature on whether these values should be included or not, and further work is planned for comparing the impact of these exclusions on values. Understanding whether respondents are not engaged or finding the task cognitively challenging is difficult to distinguish. Work on the test re-test reliability of estimates is ongoing.

Fourth, this study was conducted via the web - predominantly for cost reasons. The TTO is typically conducted in a face to face interview and while the web version of the TTO used in this study replicates the instructions and visual props of the interview, previous research indicates that the mode of administration will impact values. A potential benefit to a web administration is lack of interviewer bias and given the logical ordering of health states, it seems that the web is a feasible means to eliciting values, however further study comparing the difference in administration is warranted.

Finally, while respondents were broadly generalisable to the Canadian general population in terms of age, gender and education, since they are members of a market research panel, their generalisability is questionable. Given the objective of this study, the fact that respondents might be interview savvy, computer and internet users and english language speakers, are not primary concerns but it is plausible that given their experience, the tests for cognition might not be generalisable to the general population.

In summary, this study demonstrates DCE's as a feasible technique for valuing health states. Further work is required to inform on the benefits and disadvantages of DCE's over conventional techniques.

6. Future work

1. Test retest reliability of each method
2. Explore heterogeneity further – mixed logit and latent class models
3. Compute and compare standard errors for both sets of estimated values and calculate the number of respondents/tasks required for similar variance between approaches.
4. Improve study design
5. Extend to other countries and instruments

References

1. Patrick DL, Bush JW, Chen MM. Methods for measuring levels of well-being for a health status index. Health services research. 1973 ;8(3):228.
2. Torrance GW. Measurement of health state utilities for economic appraisal. Journal of Health Economics. 1986 Mar ;5(1):1-30.
3. Drummond MF. Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press; 2005.
4. Salomon JA, Murray CJL. A multi-method approach to measuring health-state valuations. Health Econ. 2004 Mar ;13(3):281-90.
5. Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. Health Econ. 2002 Jul ;11(5):447-56.
6. Stewart N, Chater N, Stott HP, Reimers S. Prospect relativity: how choice options influence decision under risk. J Exp Psychol Gen. 2003 Mar ;132(1):23-46.
7. Ratcliffe J, Brazier J, Tsuchiya A, Symonds T, Brown M. 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. 2009. In press.
8. Flynn TN, Louviere JJ, Marley AA, Coast J, Peters TJ. Rescaling quality of life values from discrete choice experiments for use as QALYs: a cautionary tale. Popul Health Metr. 2008 ;66.
9. McCabe C, Brazier J, Gilks P, Tsuchiya A, Roberts J, O'Hagan A, et al. Using rank data to estimate health state utility models. J Health Econ. 2006 May ;25(3):418-31.
10. Krabbe PFM. Thurstone scaling as a measurement method to quantify subjective health outcomes. Med Care. 2008 Apr ;46(4):357-65.
11. Salomon J. Reconsidering the use of rankings in the valuation of health states: a model for estimating cardinal values from ordinal data. Popul Health Metr. 2003 Dec 19;1(1):12.
12. Craig BM, Busschbach JJ, Salomon JA. Keep it simple: Ranking health states yields values similar to cardinal measurement approaches. Journal of Clinical Epidemiology. 2009 Mar ;62(3):296-305.
13. Louviere JJ, Hensher DA, Swait JD. Stated Choice Methods: Analysis and Applications. Cambridge University Press; 2000.
14. Bryan S, Dolan P. Discrete choice experiments in health economics. The European Journal of Health Economics. 2004 ;5(3):199-202.
15. Tsuchiya A, Brazier JE. Using discrete choice experiments to estimate Quality adjusted life years: the issues and methodological approaches. IHEA 2005.
16. Cairns J, Pol MVD, Lloyd A. Decision making heuristics and the elicitation of preferences: being fast and frugal about the future. Health Economics. 2002 ;11(7):655-658.
17. Lloyd AJ. Threats to the estimation of benefit: are preference elicitation methods accurate? Health Econ. 2003 May ;12(5):393-402.
18. Guttmann R, Castle R, Fiebig DG. Use of Discrete Choice Experiments in health economics: An update of the literature. CHERE working paper. 2009 ;
19. Ryan M, Gerard K. Using discrete choice experiments to value health care: current practice and future prospects. Applied Health Economics and Policy Analysis. 2003 ;255-64.
20. Ryan M, Netten A, Skatun D, Smith P. Using discrete choice experiments to estimate a preference-based measure of outcome--An application to social care for older people. Journal of Health Economics. 2006 Sep ;25(5):927-944.
21. Burr JM, Kilonzo M, Vale L, Ryan M. Developing a preference-based Glaucoma Utility Index using a discrete choice experiment. Optom Vis Sci. 2007 Aug ;84(8):797-808.
22. Brazier J, Rowan D, Yang, Y, Tsuchiya A. Using Rank and Discrete Choice Data to Estimate Health State Utility Values on the QALY scale. IHEA 2005
23. Coast J, Flynn TN, Natarajan L, Sproston K, Lewis J, Louviere JJ, et al. Valuing the ICECAP capability index for older people. Soc Sci Med. 2008 Sep ;67(5):874-82.
24. Viney R, Savage E, King M, Hossain I. Using Choice Experiments to Estimate QALYs: An Application to Prostate Cancer. IHEA 2007
25. EuroQol G. EuroQol-a new facility for the measurement of health related quality of life. Health Policy. 1990 ;16199-208.
26. Williams A. The measurement and valuation of health: a chronicle. Centre for Health Economics Discussion paper. 1995 ;136
27.Dolan P. Modelling valuations for health states: the effect of duration. Health Policy. 1996 Dec ;38(3):189-203.
28. Tsuchiya A, Brazier J, Roberts J. Comparison of valuation methods used to generate the EQ-5D and the SF-6D value sets. J Health Econ. 2006 Mar ;25(2):334-46.
29. Gudex C. Time Trade-off User Manual: Props and Self-completion Methods. 1994.

30. Shaw JW, PharmD MPH, Johnson JA, Coons SJ. US Valuation of the EQ-5D Health States: Development and Testing of the D1 Valuation Model. Medical Care. 2005 ;43(3):203.
31. Luce RD. Individual Choice Behavior: A Theoretical Analysis. New York: John Wiley & Sons, Inc; 1959.
32. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Frontiers in Econometrics. Academic Press, New York; 1974.
33. Lamers LM, McDonnell J, Stalmeier PFM, Krabbe PFM, Busschbach JJV. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ. 2006 Oct ;15(10):1121-1132.
34. Street DJ, Burgess L. The Construction of Optimal Stated Choice Experiments: Theory and Methods. 1st ed. Wiley-Interscience; 2007.
35. Huber J, Zwerina K. The Importance of Utility Balance in Efficient Choice Set Designs. Journal of Marketing Research. 1996 ;33307-317.
36. Kuhfeld WF. Orthogonal arrays. Technical report, SAS Institute. Available at http://support. sas. com/techsup/technote/ts723. html; 2006.
37. Brazier J, Ratcliffe J, Salomon J, Tsuchiya A. Measuring and Valuing Health Benefits for Economic Evaluation. 1st ed. Oxford University Press; 2007.
38. Lamers LM, Stalmeier PFM, Krabbe PFM, Busschbach JJV. Inconsistencies in TTO and VAS values for EQ-5D health states. Med Decis Making. 2006 Apr ;26(2):173-181.
39. Johnson JA, Pickard AS. Comparison of the EQ-5D and SF-12 health surveys in a general population survey in Alberta, Canada. Med Care. 2000 Jan ;38(1):115-121.

Table 1. Number of respondents and data problems (total n=691)

	TTO	**DCE**	**Both**
Respondents completed all questions (% of total)	571 (83)	588 (85)	560 (81)
1. Failed rationality check (% of completes)	30 (5)	15 (3)	8 (1)
2. All responses the same (%)	114 (20)	34 (6)	22 (4)
3. All health states given negative TTO value (%)	161 (28)	-	161 (29)
4. More than 3 logical inconsistencies in TTO states (%)	123 (22)	-	123 (22)
5. Not sufficient time spent on question (%)	-	88 (15)	88 (16)
6. No problems (%)	375 (66)	551 (94)	349 (62)*
7. 1 problem or less (%)	496 (87)	585 (99)	468 (84)
8. 2 problems or less (%)	563 (99)	588 (100)	523 (93)
9. 3 problems or less (%)	571 (100)	588 (100)	560 (100)

Notes:
1. For TTO this is initial question of whether they would prefer 10 years in full health to 10 years in impaired health. For DCE, this is a dominated choice set that all respondents see.
2 For TTO this is if all values are the same, for DCE it is if all 8 choices were the left health state or the all the right.
3. If all values for the TTO are negative
4. If there are more than three occurrences where states which are equal to better in all attributes are given lower values
5. The minimum time to load and answer the question for the DCE was determined to be 6 seconds. If questions were answered in under 8 seconds it was considered that the choice sets were not properly considered.
* denotes the respondents used in the analysis in this paper

Table 2. Characteristics of the Sample

Characteristic	**Group I Valuation sample (n=349)**	**Group II Data problems* (n=211)**	**Group III Non completes (n=131)**	**P Value†**	
				I vs II	**I vs III**
Age, mean (SE)	54 (0.77)	57 (0.91)	59 (1.08)	0.003	<0.001
Sample range	18-99	21-84	30-85		
Sex, % (n)					
Male	152 (44)	104 (50)	63 (48)	0.154	0.373
Female	197 (56)	105 (50)	68 (52)		
Highest level of education, n (%)					
Primary school or less	1 (0)	1 (1)	0	0.213	0.510
High School	57 (16)	28 (13)	20 (15)		
Community college	125 (36)	92 (44)	58 (44)		
Undergraduate degree	121 (35)	71 (34)	39 (30)		
Graduate degree	45 (13)	17 (8)	14 (11)		
Income, n (%)					
$10,000 or less	1 (0)	8 (4)	6 (5)	0.035	0.005
$10,000-$20,000	14 (4)	10 (5)	6 (5)		
$20,000-$30,000	21 (6)	10 (5)	12 (9)		
$30,000-$40,000	38 (11)	33 (16)	23 (18)		
$40,000-$60,000	111 (32)	60 (29)	36 (28)		
$60,000-$80,000	63 (18)	39 (19)	21 (16)		
$80,000-$100,000	53 (15)	26 (12)	14 (11)		
$100,000 or more	48 (14)	23 (11)	13 (10)		
Marital status, n (%)					
Married	194 (56)	117 (56)	69 (53)	0.214	0.691
Living with partner	25 (7)	14 (7)	14 (11)		
Widowed	23 (7)	21 (10)	11 (8)		
Divorced	41 (12)	21 (10)	14 (11)		
Separated	11 (3)	7 (3)	6 (5)		
Single	54 (15)	28 (13)	16 (12)		
Not reported	1 (0)	1 (1)	1 (1)		
EQ-5D dimension, n (%)					
Mobility					
Problems	78 (22)	48 (23)	25 (22)	0.913	0.961
No problems	271 (78)	163 (77)	88 (78)		
Self-care					
Problems	12 (3)	10 (5)	5 (4)	0.443	0.622
No problems	337 (97)	201 (95)	108 (96)		
Usual activities					
Problems	71 (20)	40 (19)	22 (19)	0.690	0.840
No problems	278 (80)	171 (81)	91 (81)		
Pain/discomfort					
Problems	185 (53)	115 (55)	62 (55)	0.731	0.731
No problems	164 (47)	96 (45)	51 (45)		
Anxiety/depression					
Problems	96 (28)	46 (22)	35 (31)		0.477
No problems	253 (72)	165 (78)	78 (69)		
EQ-5D UK index, mean (SE)	0.814 (0.01)	0.804 (0.02)	0.805 (0.01)	0.613	0.715
EQ-5D US index, mean (SE)	0.858 (0.01)	0.851 (0.01)	0.848 (0.02)	0.622	0.503

† T-test for continuous data, chi-square test for categorical data
* Yet to get demographic data on 2 respondents

Table 3: Parameter Estimates for DCE model

Variable	
MO2	0.118 (0.255)
MO3	-0.355 (0.300)
SC2	-0.002 (0.248)
SC3	-0.36 (0.283)
UA2	-0.308 (0.290)
UA3	-0.226 (0.276)
PD2	-0.429 (0.253)
PD3	-0.404 (0.253)
AD2	-0.347 (0.247)
AD3	-0.482 (0.283)
LY	0.841 (0.146)*
MO2*LY	-0.084 (0.102)
MO3*LY	-0.417 (0.116)*
SC2*LY	-0.087 (0.098)
SC3*LY	-0.291 (0.106)*
UA2*LY	0.052 (0.115)
UA3*LY	-0.176 (0.108)
PD2*LY	0.039 (0.102)
PD3*LY	-0.356 (0.098)*
AD2*LY	0.003 (0.095)
AD3*LY	-0.245 (0.108)*
Number of observations	2094
Mcfadden's R-sq	0.29
Log likelihood	-1024

*p<0.05
LY=Life years

Table 4: Estimates of coefficients TTO and DCE

Variable	**TTO, mean (SE) From Random effect model**	**DCE, mean Estimated (from equation 9)**
Intercept	0.138 (0.031)*	-
MO2	0.019 (0.027)	0.085
MO3	0.286 (0.032)*	0.538
SC2	0.112 (0.027)*	0.103
SC3	0.253 (0.029)*	0.389
UA2	0.110 (0.029)*	-0.025
UA3	0.112 (0.033)*	0.236
PD2	0.029 (0.028)	0.005
PD3	0.255 (0.027)*	0.471
AD2	0.066 (0.029)*	0.037
AD3	0.258 (0.028)*	0.349
Number of observations	1745	
AIC	2053	
R-sq	0.212	
RMSE	0.557	

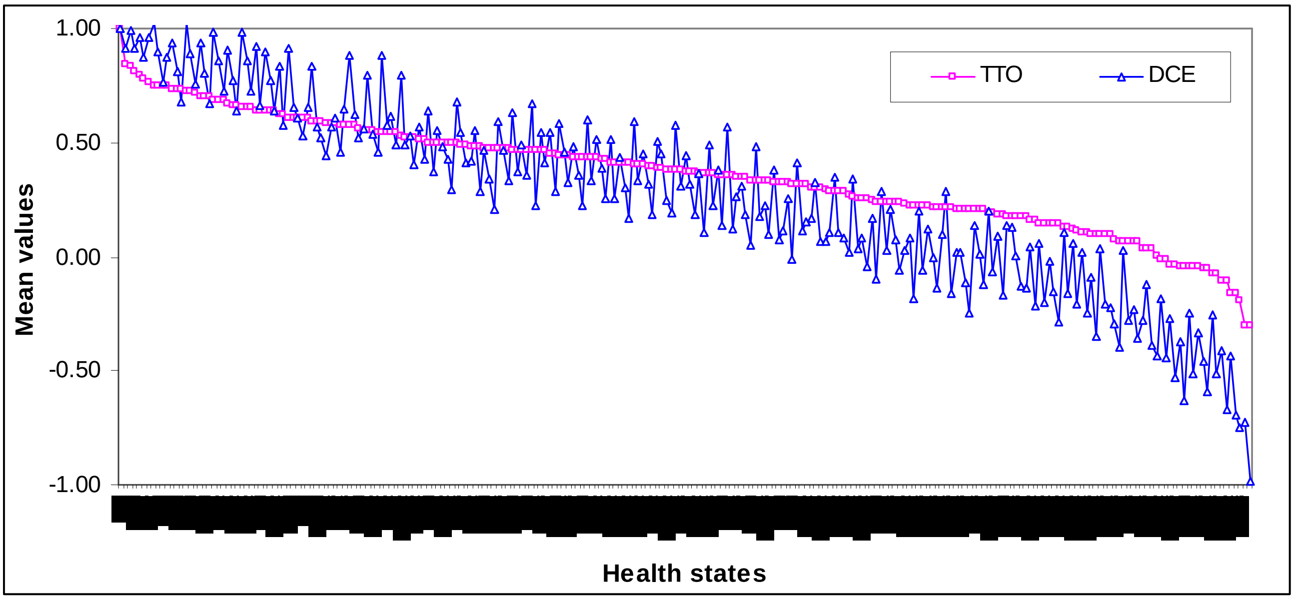
* p<0.05

Table 5. Comparison in responses between methods

	Group I Valuation sample			Group II Data problems			Group I vs Group II	
	TTO	**DCE**	**P-values†**	**TTO**	**DCE**	**P-values†**	**TTO P-values**	**DCE P-values**
Self report								
Difficulty in 'understanding', n (%)								
Very difficult	4 (1)	6 (2)	0.873	7 (3)	4 (2)	0.504	0.007	0.372
Fairly difficult	41 (12)	44 (13)		43 (21)	38 (18)			
Not very difficult	174 (50)	175 (50)		87 (41)	101 (48)			
Not at all difficult	130 (37)	123 (35)		73 (35)	68 (32)			
Difficulty in 'answering', n (%)								
Very difficult	33 (10)	45 (13)	0.398	20 (10)	15 (7)	0.850	0.403	0.052
Fairly difficult	139 (40)	143 (41)		75 (36)	78 (37)			
Not very difficult	129 (37)	113 (32)		76 (36)	78 (37)			
Not at all difficult	48 (14)	48 (14)		40 (19)	40 (19)			
Measured								
Time taken per task, mins:secs (IQR)								
1st question	2:39 (1:38-3.25)	1:07 (0:38-1:31)	<0.001	2:57 (1:39-3:36)	1:04 (0:31-1:16)	<0.001	0.097	0.900
Average of subsequent q's‡	1:56 (1:13-2:33)	0:33 (0:32-1:24)	<0.001	1:44 (1:02-2:10)	0:32 (0:20-0:36)	<0.001	0.200	0.541
Total for all q's‡	8:50 (5:26-10:04)	3:40 (2:27-4:16)	<0.001	9:27 (5:23-11:11)	3:45 (2:23-4:21)	<0.001	0.151	0.829

† 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. Comparison between estimated values for the TTO and DCE



Ordered by TTO value

Appendix A: Observed TTO and predicted values for EQ-5D states

Health State	Observed			TTO Predicted			DCE Predicted			
	n	Mean	SE	Mean	SE*	Mean Error	Mean	SE*	Mean Error vs observed TTO	Mean Error vs predicted TTO
21111	32	0.843	0.042	0.843	0.015	0.00	0.91		0.07	0.07
11121	29	0.838	0.035	0.833	0.010	0.01	1.00		0.16	0.16
11112	28	0.826	0.041	0.797	0.019	0.03	0.96		0.14	0.17
12111	37	0.824	0.037	0.751	0.018	0.07	0.90		0.07	0.15
11211	31	0.823	0.046	0.752	0.017	0.07	1.03		0.20	0.27
21112	34	0.804	0.042	0.778	0.014	0.03	0.88		0.07	0.10
21212	28	0.797	0.042	0.667	0.015	0.13	0.90		0.11	0.24
11222	32	0.735	0.048	0.657	0.016	0.08	0.98		0.25	0.33
22112	31	0.700	0.049	0.666	0.018	0.03	0.77		0.07	0.11
21321	25	0.678	0.049	0.702	0.017	-0.02	0.67		0.00	-0.03
11113	31	0.674	0.049	0.604	0.014	0.07	0.65		-0.02	0.05
11312	29	0.670	0.054	0.685	0.025	-0.02	0.73		0.06	0.04
12321	29	0.591	0.050	0.610	0.018	-0.02	0.66		0.07	0.05
23121	27	0.575	0.068	0.562	0.027	0.01	0.52		-0.05	-0.04
22222	37	0.543	0.056	0.527	0.013	0.02	0.80		0.25	0.27
13122	23	0.538	0.059	0.515	0.020	0.02	0.57		0.03	0.05
11131	23	0.537	0.075	0.607	0.017	-0.07	0.53		-0.01	-0.08
22131	29	0.506	0.068	0.477	0.013	0.03	0.34		-0.17	-0.14
31221	31	0.472	0.081	0.437	0.014	0.03	0.48		0.01	0.04
12231	34	0.467	0.054	0.386	0.019	0.08	0.45		-0.02	0.07
31211	27	0.450	0.088	0.466	0.018	-0.02	0.49		0.04	0.02
22123	31	0.429	0.085	0.445	0.017	-0.02	0.46		0.03	0.01
23311	23	0.424	0.096	0.479	0.010	-0.05	0.29		-0.13	-0.19
13311	31	0.386	0.072	0.498	0.023	-0.11	0.37		-0.01	-0.12
23222	23	0.370	0.104	0.386	0.027	-0.02	0.51		0.14	0.12
12332	27	0.368	0.083	0.318	0.015	0.05	0.15		-0.22	-0.17
22313	37	0.346	0.081	0.362	0.028	-0.02	0.23		-0.12	-0.14
13113	25	0.345	0.094	0.352	0.026	-0.01	0.26		-0.08	-0.09
21232	23	0.331	0.095	0.412	0.013	-0.08	0.43		0.10	0.02
32211	23	0.309	0.085	0.355	0.027	-0.05	0.38		0.07	0.03
12213	32	0.307	0.084	0.383	0.026	-0.08	0.57		0.27	0.19
22232	25	0.305	0.084	0.301	0.019	0.00	0.33		0.02	0.03
21333	23	0.252	0.094	0.218	0.018	0.03	-0.14		-0.39	-0.36
13223	29	0.240	0.082	0.212	0.020	0.03	0.28		0.04	0.07
21233	29	0.207	0.098	0.220	0.023	-0.01	0.12		-0.09	-0.10
23231	28	0.187	0.099	0.225	0.023	-0.04	0.08		-0.11	-0.15
11133	37	0.153	0.098	0.349	0.014	-0.20	0.18		0.03	-0.17
31323	34	0.132	0.091	0.178	0.028	-0.05	-0.13		-0.26	-0.31
23232	34	0.102	0.083	0.160	0.024	-0.06	0.04		-0.06	-0.12
32313	23	0.041	0.109	0.095	0.024	-0.05	-0.23		-0.27	-0.32
32323	28	0.029	0.108	0.066	0.019	-0.04	-0.23		-0.26	-0.30
32223	29	0.029	0.091	0.068	0.025	-0.04	0.03		0.00	-0.04
32232	25	-0.001	0.114	0.034	0.026	-0.04	-0.12		-0.12	-0.16
23233	31	-0.022	0.101	-0.033	0.022	0.01	-0.27		-0.25	-0.24
33323	27	-0.083	0.099	-0.075	0.013	-0.01	-0.52		-0.43	-0.44
33213	29	-0.125	0.095	-0.044	0.013	-0.08	-0.25		-0.13	-0.21
33332	33	-0.203	0.060	-0.109	0.014	-0.09	-0.67		-0.47	-0.56
33333	349	-0.259	0.025	-0.301	0.011	0.04	-0.98		-0.72	-0.68
MAE						0.04			0.14	0.16

Data are rank ordered by observed mean TTO values

*Predicted SE calculated using delta method, DCE values still to be calculated

Appendix B. All predicted values

Health state	TTO	DCE	Health state	TTO	DCE	Health state	TTO	DCE
11111	1.000	1.000	21213	0.475	0.591	13331	0.243	-0.096
21111	0.843	0.915	22113	0.474	0.463	13213	0.242	0.288
11121	0.833	0.995	21313	0.473	0.330	13313	0.240	0.026
21121	0.814	0.910	13221	0.470	0.631	11233	0.239	0.206
11112	0.797	0.963	13321	0.469	0.370	12133	0.238	0.078
21112	0.778	0.877	31211	0.466	0.487	11333	0.237	-0.056
11122	0.768	0.958	32111	0.465	0.359	33122	0.229	0.031
11211	0.752	1.025	11223	0.465	0.672	23231	0.225	0.080
12111	0.751	0.897	31311	0.465	0.225	23331	0.224	-0.182
11311	0.750	0.764	12123	0.464	0.544	23213	0.222	0.202
21122	0.748	0.873	11323	0.463	0.410	23313	0.221	-0.059
21211	0.733	0.940	23221	0.451	0.546	21233	0.220	0.121
22111	0.732	0.812	23321	0.450	0.284	22133	0.219	-0.008
21311	0.731	0.678	21223	0.446	0.587	21333	0.218	-0.141
11221	0.723	1.021	22123	0.445	0.458	33211	0.214	0.098
12121	0.722	0.892	21323	0.444	0.325	13223	0.212	0.283
11321	0.721	0.759	31221	0.437	0.482	33311	0.212	-0.164
21221	0.704	0.935	32121	0.436	0.354	31231	0.211	0.016
22121	0.703	0.807	31321	0.436	0.221	13323	0.211	0.021
21321	0.702	0.674	13212	0.434	0.599	32131	0.210	-0.112
11212	0.687	0.988	13312	0.432	0.337	31331	0.210	-0.246
12112	0.685	0.860	11232	0.431	0.517	31213	0.208	0.138
11312	0.685	0.726	12132	0.430	0.389	32113	0.207	0.010
21212	0.667	0.903	11332	0.430	0.256	31313	0.207	-0.123
22112	0.666	0.774	23212	0.415	0.514	23223	0.193	0.198
21312	0.666	0.641	23312	0.413	0.252	23323	0.192	-0.064
11222	0.657	0.983	21232	0.412	0.432	33221	0.185	0.093
12122	0.656	0.855	22132	0.411	0.303	33321	0.183	-0.169
11322	0.656	0.722	21332	0.411	0.170	31223	0.179	0.134
12211	0.641	0.922	13222	0.405	0.594	13232	0.179	0.128
12311	0.639	0.660	13322	0.403	0.333	32123	0.178	0.005
21222	0.638	0.898	31212	0.401	0.450	31323	0.178	-0.128
22122	0.637	0.770	32112	0.399	0.321	13332	0.177	-0.134
21322	0.637	0.636	31312	0.399	0.188	23232	0.160	0.043
22211	0.622	0.837	23222	0.386	0.509	23332	0.158	-0.219
22311	0.620	0.575	12231	0.386	0.451	33212	0.148	0.061
12221	0.611	0.918	23322	0.384	0.247	33312	0.146	-0.201
12321	0.610	0.656	12331	0.384	0.190	31232	0.146	-0.021
13111	0.610	0.611	12213	0.383	0.574	32132	0.144	-0.150
11131	0.607	0.529	12313	0.381	0.312	31332	0.144	-0.283
11113	0.604	0.651	31222	0.372	0.445	12233	0.127	0.103
22221	0.592	0.832	32122	0.370	0.317	12333	0.126	-0.159
22321	0.591	0.570	31322	0.370	0.183	33222	0.119	0.056
23111	0.591	0.525	22231	0.366	0.366	33322	0.117	-0.206
21131	0.588	0.444	22331	0.365	0.104	22233	0.108	0.018
21113	0.585	0.566	22213	0.364	0.488	22333	0.107	-0.244
13121	0.581	0.606	22313	0.362	0.227	32231	0.100	-0.087
31111	0.577	0.462	32211	0.355	0.384	32331	0.098	-0.349
11123	0.575	0.647	13131	0.355	0.140	32213	0.097	0.035
12212	0.575	0.885	12223	0.353	0.569	13133	0.097	-0.209
12312	0.573	0.623	32311	0.353	0.122	32313	0.095	-0.226
23121	0.562	0.521	13113	0.352	0.262	23133	0.078	-0.294
21123	0.556	0.561	12323	0.352	0.307	33131	0.069	-0.398
22212	0.556	0.800	11133	0.349	0.181	32223	0.068	0.031
22312	0.554	0.538	23131	0.336	0.055	33113	0.066	-0.276
31121	0.547	0.457	22223	0.334	0.484	32323	0.066	-0.231
12222	0.546	0.880	23113	0.333	0.177	31133	0.063	-0.358
13112	0.544	0.574	22323	0.333	0.222	33123	0.037	-0.281
12322	0.544	0.619	21133	0.330	0.095	32232	0.034	-0.124
11132	0.542	0.492	32221	0.326	0.379	32332	0.032	-0.386
22222	0.527	0.795	33111	0.324	0.072	33132	0.003	-0.436
23112	0.525	0.488	32321	0.324	0.117	13233	-0.014	-0.183
22322	0.525	0.533	13123	0.323	0.258	13333	-0.015	-0.445
21132	0.523	0.407	31131	0.322	-0.009	23233	-0.033	-0.269
13122	0.515	0.569	12232	0.320	0.414	23333	-0.034	-0.530
31112	0.511	0.424	31113	0.319	0.113	33231	-0.041	-0.373
13211	0.500	0.636	12332	0.318	0.152	33331	-0.043	-0.635
13311	0.498	0.374	23123	0.303	0.172	33213	-0.044	-0.251
11231	0.497	0.554	22232	0.301	0.329	33313	-0.046	-0.513
23122	0.496	0.484	22332	0.299	0.067	31233	-0.047	-0.332
12131	0.496	0.426	33121	0.295	0.068	32133	-0.048	-0.461
11331	0.495	0.293	31123	0.289	0.108	31333	-0.048	-0.594
11213	0.494	0.677	32212	0.289	0.347	33223	-0.073	-0.255
12113	0.493	0.548	13132	0.289	0.103	33323	-0.075	-0.517
11313	0.492	0.415	32312	0.288	0.085	33232	-0.107	-0.410
31122	0.482	0.420	23132	0.270	0.017	33332	-0.109	-0.672
23211	0.481	0.551	32222	0.260	0.342	32233	-0.158	-0.436
23311	0.479	0.289	33112	0.258	0.035	32333	-0.160	-0.697
21231	0.478	0.469	32322	0.258	0.080	33133	-0.189	-0.747
22131	0.477	0.341	31132	0.256	-0.046	33233	-0.299	-0.722
21331	0.476	0.207	13231	0.244	0.165	33333	-0.301	-0.983