Title: A Methodological Exploration into the Feasibility of Eliciting Preference Values for the new 7 Item Weight-specific Adolescent Instrument for Economic evaluation (WAItE) using Discrete Choice Experiments (DCEs)

Authors

PhD student: Yemi Oluboyede[1]

Supervisors & specialist support: Claire Hulme[1], Andrew Hill[1] Aki Tsuchiya[2] & Stephane Hess[1]

[1]University of Leeds, [2]University of Sheffield

Corresponding author:

Yemi Oluboyede

Academic Unit of Health Economics

Leeds Institute of Health Sciences

University of Leeds

101 Clarendon Road

Charles Thackrah Building

Leeds LS2 9LJ

Tel: +44 (0)113 343 0823

E-mail: y.oluboyede@leeds.ac.uk

Project website: http://www.leeds.ac.uk/lihs/research/AUHE/QoL-obesity.html

ACKNOWLEDGEMENTS

We would like to acknowledge the advice and support of the following individuals: Cathy Brennan, Jenny Hewison, Donna Lamping, Christopher McCabe, David Meads, Jennifer Roberts, Katherine Stevens, and Alan Tennant and members of Academic Unit of Health Economics, University of Leeds. A huge thanks to Survey Sampling International (SSI) and the survey participants. The usual disclaimer applies.

The work presented here is part of a National Institute for Health Research (NIHR) funded fellowship project (Developing a preference based quality of life measure for use in economic evaluation of weight management interventions aimed at the young). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

INTRODUCTION

In England, 23.6% of men and 24.4% of women over the age of 15 are obese and a further 41.4% and 32.0% overweight (Health Survey for England, 2008). In addition 20% of boys and 16.5% of girls (average 18.3%) aged 10-11 years are classified as obese (The Health and Social Care Information Centre, 2009). Due to the long term health risks of childhood obesity (such as respiratory diseases and type 2 diabetes) and the emotional and psychological effects of being obese, public health initiatives and specialised weight management interventions have been developed specifically for the adolescent population (NICE Guideline, 2006).

Economic evaluations are used to inform resource allocation decision making within the NHS (NICE, 2008). Economic evaluations typically assess the incremental costs and benefits associated with a new intervention, compared with standard practice. Economic evaluation can be used to inform resource allocation decision making for weight management interventions. Within England and Wales the National Institute for Health and Clinical Excellence (NICE) methodology guidelines are used for the appraisal of new and existing technologies (NICE, 2008). The guidance states that the quality-adjusted-life-year (QALYs), which combines both length and quality of life (consisting of anchor points at zero (the 'dead' state) and one (full health)) into a single summary measure, should be used for the assessment of benefits. In order to calculate QALYs a valid and reliable tool to measure and assign values to different aspects of weight specific QoL is needed.

NICE identifies the EQ-5D as the appropriate source for obtaining the utility/preference weights for the QALY (NICE, 2008). The applicability of the EQ-5D is however limited in the assessment of weight management interventions as it may not fully reflect the affect of weight status or changes in weight status on quality of life (QoL). For example important domains influenced by weight status such as 'public distress' and 'sexual life' are not fully captured in the EQ-5D (Brazier et al., 2004), whilst "condition specific" instruments may be more applicable for the assessment of weight status and changes in weight status on QoL (Kolotkin et al., 2006). Although generic QoL and health related QoL (HRQoL) instruments created specifically for the adolescent population, are more likely to be relevant to the younger population, they have been found to perform poorly in discriminating between different body-mass index (BMI) subgroups on physical, psychological and emotional dimensions of functioning relative to weight specific measures (Kolotkin et al., 2006). The value in using weight specific QoL measures to assess interventions for overweight and obese adolescents is therefore strengthened. Three non-preference based self report weight specific instruments targeting 11 to 18 year olds were identified in the literature (Kolotkin et al., 2006, Zeller & Modi, 2009 and Morales et al., 2011). However, as the existing weight specific QoL measures lack preference weights, they cannot be used to calculate QALYs and are therefore not suitable for cost utility analysis (CUA).

The current study builds on from previous work that undertook the creation of a new measure, suitable for the elicitation of weight specific preference values, the seven item Weight-specific Adolescent Instrument for Economic evaluation (WAItE). The content of the WAItE was informed by undertaking qualitative interviews with adolescents aged 11 to 18 years. Psychometric assessments and Rasch analysis were then carried out to refine the instrument. WAItE has seven dimensions each with five levels. This study reports on on-going work exploring the different methodological applications of discrete choice experiments (DCEs) in the creation of a robust value set for the WAItE. The current study reports on a methodological investigation to assign preference values to weight specific QoL states described by the WAItE so that it can be used to carry out fully informed CUA of weight management interventions aimed at the adolescent population.

Elicitation of preference values in health economics

Preference based measures have been developed from economics and decision theory in order to provide an estimate of individuals' preferences for different health states (Drummond, 1996; Bakker & van der Linden, 1995). This form of instrument uses preference-based methods to value health states. These states can be defined with the use of a descriptive system. Descriptive systems are made up of domains, items within domains and response categories for each item. Health states are valued by individuals using preference elicitation techniques to derive preference values (or utility weights) for each health state. For a more detailed description of elicitation techniques available for valuation of health states; and the individuals who can be asked to value health states see Brazier et al. (2007). Recent studies have shown that discrete choice experiments (DCEs), a quantitative technique for eliciting preferences that requires respondents to state their choice over sets of hypothetical alternatives or scenarios, are emerging as another alternative to Standard Gamble (SG) and Time Trade-Off (TTO) (Lancester and Louvier, 2008). Although DCEs are more complex to design and analyse, they are easier to implement without the presence of an interviewer and can be administered using an on-line survey. This means that a large amount of data can be collected over a short period of time.

An overview of the discrete choice experiment (DCE) technique

DCEs draw upon Lancaster's economic theory of value (Lancaster, 1966), random utility theory (McFadden (1973, pp. 105-42)) and involve three main inter-related components: a). An experimental design used to implement a choice survey and generate choice data, b). Discrete choice analysis to estimate preferences from choice data; and c). Use of the resulting model to either derive welfare measures or construct other policy analyses (Lancester & Louvier, 2008). Random utility theory is derived from the economic assumption that a rational individual will select the option from a set of alternatives that provides them with the greatest expected utility. The first step in formulating the econometric model is to specify an indirect utility function that relates to attribute levels and individual characteristics to the level of utility enjoyed. The decision making process within a DCE can be seen as a comparison of indirect utility functions. The indirect utility function used in analysis is merely an approximation to the individual's actual indirect utility function. It is assumed that a respondent makes a series of choices and for each choice an alternative that leads to the higher level of utility is picked by the respondent. An econometric representation of the decision is illustrated below (as per Ryan and Gerard (2003 pp. 55)).

$$U_{iq}\ (A) = \ vi_q\ (A) + \ \varepsilon_{iq} \tag{1}$$

Here $U_{iq}\ (A)$ represents the indirect utility function of individual *q* for alternative *i* with attributes *A*, $vi_q\ (A)$ represents the measurable component of utility estimated empirically. The utility predicted by this approximation will almost certainly not be 100% correct. To complete the econometric model, a random element is added, ε_{iq} which reflects the unobservable factors, affecting the estimation of the indirect utility function. This error component picks up the difference between the true utility derived from choosing a particular option from a pre-defined set of alternatives and that which is modelled. The inclusion of this random element explains the derivation of the title random utility model (RUM). The respondent will choose alternative *i* over alternative *j* if:

$$\left(v_{iq} + \ \varepsilon_{iq}\right) > \left(v_{jq} + \ \varepsilon_{jq}\right) \tag{2}$$

or

$$(v_{iq} - v_{jq}) > (\varepsilon_{jq} - \varepsilon_{iq}) \quad (3)$$

Since there is an error part of the utility function, the analysis becomes one of probabilistic choice. The probability that any particular respondent prefers a particular option to any other alternative available to them, can be expressed as the probability that the utility associated with that option, according to the model, exceeds that associated with all other options available to them. Given that the error terms are unknown, a probability model is estimated where:

$$P(i_q|A,C) = Pi_q = [(v_{jq} - v_{iq}) > (\varepsilon_{iq} - \varepsilon_{jq})] \quad (4)$$

The probability of choosing alternative i (over j) by individual q, given the set of attributes A and the choice set C, is given by the probability that the error difference is smaller than the difference of the 'observable' utility component between i and j (Ryan and Gerard, 2003).

DCEs are regularly used in health economics to elicit preferences over healthcare products, programmes and in the valuation of preference for health states (Ryan et al., 2006, Viney et al., 2007, Coast et al., 2008 and Flynn et al., 2008). Several reviews have been conducted to compile the evidence of the use of DCEs in the health care context (Ryan et al., 2001a&b, 2003 and de Bekker-Grob et al., 2012). Studies have only recently endeavoured to anchor the estimated utility values on the full health – dead scale, using the DCE_{TTO} technique, where QALY calculations can be made (Bansback et al., 2012 and Norman et al., 2012). This is done by incorporating an additional 'years of survival' (duration) attribute to the design of a DCE and then creating a multiplicative design that allows for interactions between each of the QoL attribute levels and the life years attribute (which is treated as a continuous variable).

The proposed research

In the current methodological study, both the novel DCE_{TTO} valuation technique proposed by Bansback et al. (2012) with duration as an attribute and DCE with no duration, anchored using two strategies based on additional data ($DCE_{NoDuration}$), are assessed. The overall aim of this study was to explore alternative applications of DCEs in valuing states described by the WAItE. As DCE_{TTO} has never been tested utilising a mild descriptive system, there is a risk that people might not be willing to trade duration in exchange for better QoL, and so $DCE_{NoDuration}$ was also implemented. Assessments of the sign and significance of the parameter coefficients generated from the alternative models were undertaken in order to identify the optimal valuation method and to test the following:

DCE_{TTO} & $\text{DCE}_{\text{NoDuration}}$	• Can respondents distinguish between different levels of the WAItE dimensions? • Do respondents self report understanding the DCE_{TTO} and $\text{DCE}_{\text{NoDuration}}$ tasks?
DCE_{TTO}	• Do longer duration options of have an impact on DCE_{TTO} coefficient estimates? • Does respondent age have an impact on DCE_{TTO} coefficient estimates?

The empirical assessment of the aforementioned issues will provide evidence to support or reject the use of DCE_{TTO} or $\text{DCE}_{\text{NoDuration}}$ in the valuation of the WAItE. The structure of the paper is as follows: The methods used in testing the aforementioned questions are reported, the DCE analysis is described, and this is followed by the presentation of the results. The paper ends with discussion around the findings and conclusions.

METHODS

The general methods used in the design of this study were informed by the DCE user's guide developed by Lancester and Louvier (2008). The methods specifically employed in the data analysis were informed by the Bansback et al. (2012) study described above. The data utilised in the paper was generated from three pilot studies, sample size calculations were not undertaken as the information generated from the pilot studies were to inform the main valuation study to follow.

Survey and valuation tasks

A web survey was conducted asking respondents to complete a series of DCE pairwise tasks made up of two DCE profiles (two alternatives). Each of the DCE profiles were made up of a health state defined by the WAItE, containing seven attributes (henceforth referred to as the QoL attributes), with five possible levels for each attribute. In this case level 1 referred to the best level in each attribute 'never' – so health state 1111111 is referred to full health – and 5555555 is referred to the worst health state possible in this descriptive system. For the DCE_{TTO}, in addition to the seven QoL attributes, an eighth attribute describing the number of years an individual would live in a particular state followed by immediate death, the 'duration' attribute, was added making up the full DCE_{TTO} profile. The duration attribute was also made up of five levels of life years. Three different duration versions were used: baseline: 10, 9, 8, 6 & 5, and then these were scaled up by a factor of two (i.e. 20, 18, 16, 12, and 10 yrs) and five (i.e. 50, 45, 40, 30 and 25 yrs) to assess the impact of two sets of longer duration options on coefficient estimates (Table 1). As different duration levels were utilised and, which may be influenced by time preference rates, a time preference question was also added to the survey (Figure 1). Respondents were asked whether they would prefer to have ill health now followed by good health in the future (negative time preference) or good health now followed by ill health in the future (positive time preference), or whether they had no preference (zero time preference).

On entering the survey, at the informed consent stage, respondents were told: *You are invited to take part in a research study in the area of weight status and how it can affect different aspects of* *life...We would like to find out which of the different aspects of life that can be affected by weight status, are most important to you*, so that they knew the subject matter of the survey. Respondents first completed questions on background characteristics, socioeconomic status, self-assessed health and also the WAItE. Self report weight and height data was also collected so that BMI could be calculated. The survey presented respondents with information about the DCE tasks and they were asked to complete a practice question before completion of the main DCE section of the survey. The DCE tasks were made up of two profiles, and respondents were asked to choose which alternative they preferred. Here, respondents are forced to make a choice between the two alternative profiles - an opt-out was not available. Respondents completed 10 DCE_{TTO} or $DCE_{NoDuration}$ tasks each. The last section in the survey asked respondents about any difficulties they had in understanding and answering the DCE tasks. An example of the choice set included in the survey is given in Figure 2.

Recruitment & Sample

A market research company, with an existing internet panel, was commissioned to manage and host the website were the survey was held. Potential participants were recruited from a panel of individuals who subscribe to the company to complete questionnaires in exchange for monetary incentives on completion of each questionnaire. The company recruited a sample of the UK population over 18 years from the panel based on quotas in terms of gender and age in order to obtain a balanced sample of respondents. A randomisation coding was applied so that there was an equal response to each of the profile blocks made up of 10 DCE tasks (explained below). A speeder check was implemented, whereby participants were screened out of the study if they completed the DCE section of the survey within the minimum time limit. The minimum limit was set based on the assessment of how long it took the majority of participants (90%) from the first pilot study to complete the DCE section of the survey.

There were three independent surveys (pilot studies) completed by three different groups of participants. Pilot A –the first sample was given the DCE_{TTO} task with duration options of 10, 9, 8, 6 and 5 years (D10), similar to those used in the Bansback et al. (2012) study. Pilot B – the second sample completed the DCE_{TTO} task with the baseline duration options scaled up by a factor of two (D20) and a factor of five (D50). Pilot B was split into three: Sample 1 consisted of 18-54 year olds who were given the duration options that were scaled up by five. Sample 2 consisted of 55 plus year olds given the duration options scaled up by two. Sample 3 consisted of 35-54 year olds with duration scaled up by two. Three samples were used for Pilot B in order to test the impact of different duration level options on different age groups. The pairwise alternatives for Pilots A and B were identical with the only difference being the duration options presented to individuals. Pilot C – the final sample was presented with the $DCE_{NoDuration}$, where there was no duration attribute, but instead two different anchoring tasks. The pairwise alternatives for Pilot C was different to the ones used in Pilots A and B as the design did not include a duration attribute. The anchoring tasks consisted of a Visual Analogue Scale (VAS) where respondents were asked to value the WAItE PITS state (5555555) (Figure 3). The second anchoring task was a binary choice version of the TTO, between the PITS state for a longer duration and full health (1111111) for a shorter duration, where respondents were allocated a different combination of values.

Experimental design

For DCE_{TTO} a factorial design (experimental designs in which every level of every variable is paired with every level of every other variables under consideration) made up of the seven QoL attributes (WAItE items), duration attribute and associated levels, summarised in Table 1, results in 390625

(5^8) possible profiles. For $DCE_{NoDuration}$ the total number of possible scenarios that are made up of only the QoL descriptive system is 78125 (5^7). A full factorial design would result in some scenarios where a dominant alternative is present (i.e. where all attribute levels are better in profile A compared with profile B, for example, in a given pairwise choice set). For the WAItE, there are no implausible scenarios (where the worst level in one dimension is incompatible with the best level in another) and so this did not have to be factored into the design. An efficient factorial experimental design was constructed consisting of nine blocks each containing 10 pairwise choice tasks. The designs for the DCE_{TTO} and for the $DCE_{NoDuration}$ were generated using the Ngene design software (ChoiceMetrics, 2012) and were guided by advisors with expertise in the area of the design and analysis of SPs. The DCE_{TTO} design accounted for interactions between the WAItE dimension levels and the duration attribute, the latter was treated as a continuous variable. A row-swapping process was used based on constraints to avoid dominance. For each design, the Ngene programme was left to run for tens of thousands of iterations and the one that displayed the lowest D-error (a measure of the (in)efficiency of the experimental design) was chosen with the aim of minimising the standard errors. The reliability of the model parameters to be estimated can be quantified in terms of the asymptotic standard errors and covariances; thus improvements in reliability suggest a reduction in the asymptotic standard errors (Bliemer et al., 2008). The software generated a sample of 90 profiles to be included in the DCE survey. Each of the nine blocks of profiles was randomised in the survey so that an equal number of respondents were given each of the nine blocks, for both the DCE_{TTO} and $DCE_{NoDuration}$.

Analysis

The analysis described below for the DCE_{TTO} is based on the two step process developed by Bansback et al (2012) to generate utility values on a scale anchored on the full health dead scale using DCE_{TTO}. Step 1: Makes use of choice models to estimate the utility function based on the characteristics of the alternatives (i.e. modelling the choices as a function of the WAItE and duration level attributes). Here the multinominal logit (MNL) model is used. The model aims to find the values of parameters such that the chosen alternative always has the highest utility, thus maximizing the log-likelihood of the actual observed choices. Hakim & Pathak (1999, pp. 106) summarise the multinominal logit model as follows:

$$\mathrm{P}\ (\mathrm{x}|\ \mathrm{A}) = \ \mathrm{e}^{\mathrm{V(x)}} | \sum \mathrm{e}^{\mathrm{v(j)}} \qquad (5)$$

For all $\mathbf{j}$ in $\mathbf{A}$ where $\mathrm{P}\ (\mathrm{x}|\ \mathrm{A})$ = probability of selecting a particular alternative $\mathbf{x}$ from a given choice set $\mathbf{A}$; $\mathrm{V}(\mathrm{x})$ = systematic, observable component or mean value of the choice alternative $\mathbf{x}$; $\mathbf{e}$ = natural constant, the base of the natural logarithms = 2.7183.

To assess the sign and the magnitude of the DCE_{TTO} and $DCE_{NoDuration}$ coefficients, the analysis described in Bansback et al (2012) is followed for estimating the interaction effects between the WAItE dimension levels and the duration attribute. The utility function $\boldsymbol{\mu}$ of each respondent $\mathbf{z}$ is defined as multiplicative between the vector of levels for each WAItE attribute $\mathbf{x}$ and duration in each scenario so that:

$$\mu_{zj} = \propto + \beta t_{zj} + \gamma' x_{zj}.t_{zj} + \varepsilon_{zj} \qquad (6)$$

β = value of living in full health for the specified duration. γ = the disutility of living in the dysfunctional set of WAItE states for the same duration. ε_{zj} = is the random term (assumed to be independent and identically distributed). t = duration, which treated as continuous for the majority of the models. In addition to modelling the interaction effects between the WAItE and duration attributes. Main effects models were also estimated where the dimension level dummies and duration were not interacted with each other. These models were estimated purely in order to assess model performance in terms of investigating whether the un-interacted parameter estimates behaved as we would expect. Both the main effects and interaction models were estimated for DCE_{TTO} where different duration variants were used.

The second step of the Bansback et al (2012) approach involves the modelling/derivation of the health state values using the DCE coefficients obtained from Step 1. Bansback et al (2012) show how the value for each health state can be anchored on the full health - dead health utility scale using an anchoring utility function. As Step 2 is contingent on the coefficients from Step 1 having the correct sign and consistent ordering, the analysis in this paper focuses on the coefficient estimates derived from the multinominal logit models in Step 1.

The Bansback et al (2012) approach was specifically designed for the inclusion of a duration attribute within a DCE task. Where there is no duration attribute, equation six cannot be derived. In order to be able to directly compare DCE_{TTO} and $DCE_{NoDuration}$,an assessment of the WAItE dimension level parameter estimates is necessary. This was done by only estimating the WAItE attribute parameters with no duration parameter. For DCE_{TTO}, choice sets were included in the model where the duration levels were the same between the two profiles making up a pairwise choice set. The duration attribute was not included in the $DCE_{NoDuration}$ tasks and thus there was no duration parameter to be estimated using this method.

RESULTS

The Sample

Between the three pilot studies, a sample of 757 members of the market research panel started to complete the survey. Of these 93 (12%) were screened out because they completed the DCE section of the survey within the minimum time limit of 75 seconds (this was based data from Pilot A - 90% of participants completed the DCE section in 75 seconds or more). 155 (20%) individuals dropped-out whilst answering the non DCE sections of the survey, whilst 17 (2%) dropped out whilst answering the DCE questions. Overall a total of 489 (65%) of the participants who initially started the survey were included in the DCE analysis. This was made up of: 112 individuals in Pilot A, 206 individuals in Pilot B and 171 individuals in Pilot C (Table 2). The characteristics of the respondents are shown in Table 3. Most of the respondents were in excellent to fair health 441 (90%). The majority of participants were either overweight or obese (having a BMI of 25 and over) 275 (56%). In terms of self assessed understanding of the DCE tasks only 7 (1% of the total number of participants) felt that they did not understand what was being asked of them.

Model Results

DCE_{TTO} long vs. short durations – Main effects models

Table 4 presents the results, for the main effects models (for Pilot B where Samples 1 o 3 are combined and different duration levels have been utilised, the duration attribute is estimated as a categorical variable). In the majority of the models, the estimated WAItE dimension level dummy parameter coefficients do not have the expected sign and are not statistically significant at the 10% level (significance was assessed at the 10% level in order to take into account the small sample size of each of the pilot studies). For Pilot A, only 4 (14%) out of the 28 WAItE dimension level dummy variables have the expected negative sign. Pilot B seems to generate the better performing models, in terms of the highest number of estimated WAItE dimension level parameters with a negative sign. The best model used data from participants in Sample 3 and composed of 35-54 year olds who were presented with the duration levels scaled up by two (i.e. 20, 18, 16, 12 or 10 yrs). Here 26 (93%) of the estimated WAItE dimension level dummy coefficients have the expected negative sign and the majority of the parameters are statistically significant at the 5% level. The coefficient for the constant term is significant only in this model. A significant constant term suggests that there is specification error in the analyses (Bansback et al., 2012) and thus we would expect the coefficient for the constant term not to be significant. The next best performing model is when all the data from Pilot B is combined from the Samples 1, 2 and 3 estimating 16 (57%) of the WAItE level parameters with a negative sign. The duration coefficients for both these models are positive, indicating respondents prefer more years of life, as we would expect, and the coefficients are statistically significant.

The results from the main effects model also suggest that the presenting longer durations scaled up by more than two does not generate the parameter estimates with the sign we would expect. This is illustrated in the estimated parameter coefficients of Pilot B Sample 1, based on data obtained from 18 – 54 years olds given duration options scaled up by 5 (duration levels of 50, 45, 40, 30 and 25 yrs were presented in the DCE profiles). Here only 5 (18%) of the estimated WAItE dimension level coefficients were negative. Overall, the results from the main effects models suggest that only the 35-54 year olds can on average consistently distinguish between different levels of the WAItE dimensions, and 54 plus year olds are on average not able to distinguish between the different levels, as only 4 (14%) of the parameters estimated for the WAItE level dummy variables were negative. The duration coefficient is positive for the model based on the data from Pilot B Sample 3, though in this case it is not significant (p=0.1).

DCE_{TTO} long vs. short durations – Interaction models

Table 5 presents the models estimating the interaction effects between the WAItE dimension levels and the duration attribute. We would expect the interaction coefficients and the duration coefficient to have the opposite signs. If the interaction coefficients between the WAItE and duration attributes were negative, this would mean that, on average respondents prefer to live in longer health profiles and in less severe levels of each health status attribute. For the majority of the models, however, the attribute coefficients did not have the expected sign. Again, where data from Pilot B Sample 3 is utilised, the majority of the parameter estimates for the WAItE and duration interaction dummies have the correct sign (20 (71%)) and the next best performing model is generated when the all of the data from Pilot B is combined (17 (61%)). Nevertheless the majority of these parameters are not statistically significant at the 10 % level.

When we consider the results from Pilot B Samples 1 and 2, care needs to be taken when interpreting these results. The majority of the coefficients for Pilot B Sample 2 (generated from data composed of 55 plus year olds presented with duration options scaled up by 2) are statistically significant at the 5% level. However, contrary to what we would expect, the interaction parameter estimates have a positive sign whilst the duration attribute has a negative sign. This implies that on average, respondents preferred to live in shorter health profiles and in more severe levels of each health status attribute. However as was stated in the previous paragraph, as long as interaction coefficients and the duration coefficient to have the opposite signs, the estimated parameter coefficients can be used to calculate utility values for the WAItE. Intuitively this is questionable. However, when we consider Step 2 of the Bansback et al (2012) methodology, the anchoring utility function, utilises the coefficient estimates in Step 1 through the calculation of the ratio between interaction parameter estimates and the duration parameter estimate and in order for the function to work correctly the ratio between the duration and WAItE coefficients should be negative, meaning that *either* the numerator (the interaction coefficient) or the denominator (the duration coefficient) has to be negative.

DCE_{TTO} vs $DCE_{NoDuration}$ – Main effects models

Table 6 shows the results of comparing the DCE_{TTO} and $DCE_{NoDuration}$ methods. The data used for the DCE_{TTO} models is based on the pairs where duration is the same across the pair (although the DCE_{TTO} design will now be inefficient as this was not its original purpose). The results in Table 6 show, where there is no duration attribute, the model utilising data from Pilot C performs the best in terms of the number of estimated negative WAItE level parameters. However the design for this study was based on having no duration attribute and so it is more likely to provide more robust parameter estimates. Bearing this in mind, Pilot B Study 3, as with the two models previously discussed, also generated a high number of negative WAItE coefficients. Overall there are very few statistically significant estimates and the standard errors of the parameter estimates for Pilot B Sample 2 and 3 were very large.

$DCE_{NoDuration}$

The results for the anchoring tasks are shown in Figures 4 and 5. When the VAS was used, the values participants assigned to the PITS state covered the full length of the thermometer. The mean value for the PITS state was 0.23 and the median value was 0.20 when VAS was utilised. Comparatively, when the paiwise anchoring question was used the median value for the PITS state was 0.75.

Time preference

The responses to the time preference question included in Pilot B showed that 49% of the respondents had positive time preference 41% negative time preference and 10% had no preference.

Discussion & Conclusion

This paper presents ongoing work assessing alternative DCE methodologies for estimating health state values for a new weight specific measure – the WAItE. Although none of the model results generated consistently ordered and statistically significant estimates for the WAItE dimension level coefficients, the small sample sizes used in the pilot studies were unlikely to produce robust and definitive parameter estimates. DCE sample size estimates are generally based on rules-of-thumb and budget constraints. The minimum sample size for conjoint studies normally ranges between 150 and 1200 respondents (Orme, 1998). The data from the few individuals who self assessed themselves as not understanding the DCE task were excluded from the analysis, however this is open for debate as one of the major benefits of DCE is that it can cope with the inclusion of data from these individuals. Moreover the parameter estimates that were generated from the pilot studies provide a good indication of what to expect when a larger DCE valuation study is undertaken.

The DCE_{TTO} methodology had been shown to be successfully applied to the EQ-5D, a descriptive system for more severe health states. However, the use of longer time horizons were also assessed in the current study as the time horizons offered to respondents based on the duration levels used in the Bansback et al (2012) study could be perceived as potentially very short when combined with a comparatively mild descriptive system such as the WAItE. Two sets of longer duration options were offered in order to test whether short survival prospects *combined with* the relatively mild health states would lead respondents not to trade years of life for better health. In order to test the *pure* effect of having scaled up durations, different age groups were given different duration options. The principal findings from adopting the DCE_{TTO} in this study were as follows: If the DCE_{TTO} approach developed by Bansback et al (2012) is to be adopted for the valuation of the WAItE, the duration levels need to be carefully considered. The baseline duration levels used in conjunction with the valuation of the EQ-5D do not seem to be compatible with the milder health states described by the WAItE. Conversely, the use of extremely long duration level options of 50, 45, 40, 30 and 25 yrs for WAItE DCE profiles also seem not to produce parameter estimates that would be expected. The best performing model using DCE_{TTO} seems to be generated from data obtained from 35-54 year olds valuing DCE profiles utilising duration levels of 20, 18, 16, 12 or 10 yrs. When the same duration levels are implemented for individuals aged 55 plus years, the estimated parameter estimates generated are not as we would expect. The implications of this are that only this age group seem to be able to distinguish between the dimension levels of the WAItE. A supplementary analysis, of the data obtained from only 35 – 54 year olds from Pilot A utilising baseline duration levels of 10, 9, 8, 6 or 5 yrs and Pilot B Sample 1 where 50, 45, 40, 30 and 25 yrs duration levels were used could be undertaken to assess whether the expected parameter coefficients are estimated. If so, this would imply that 34-55 year olds can distinguish between the WAItE dimension levels regardless of the duration levels that are used.

Time preference is an important issue to consider when asking individuals to consider health benefits further in the future rather than in the immediate future. The analysis currently assumes proportionality. So a DCE_{TTO} choice involving 10years in a slightly dysfunctional WAItE state vs 8 years in a very slightly dysfunctional WAItE state should be the same as 40 years in the former state vs 32 years in the latter state (i.e. interpreted as identical choice set). If people pay more attention to the 8 year difference in the latter pair (as opposed to the 2 year difference in the former pair), ignoring that it is over 40 years instead of 10, then it means the assumption of proportionality may not hold. Also if people have a positive time preference (i.e. prefer benefits now rather than in the future), then a benefit of 8 years not starting for 32 years may be valued less than a benefit of 8 years starting next year, say. Therefore it could be argued that if longer duration options are being used, then time preference rates should also be elicited. A counter argument to this is that, as DCEs aim to elicit aggregate rather than individual level values, at the aggregate level, time preference could be assumed to be zero. This latter argument was tested empirically with the inclusion of a time preference question at the end of the survey for Pilot B. The results showed that there was approximately an even split between the sample with those individuals with a positive time preference and those with a negative time preference. If at the aggregate level, about half of the individuals included in the sample had positive or negative time preference. If this rate of time preference varies so that there are those with positive and negative time preference and in equal measure, aggregate time preference could potentially equate to zero. If this was the case then the time preference rate at the aggregate level would be zero and perhaps the elicitation of time preference rates when longer duration levels are implemented would not be necessary.

In order to address the situation where individuals may have been unwilling to trade any life to live in a better WAItE, $DCE_{NoDuration}$ was also considered. This situation would come about if making the duration levels longer was not enough to encourage trading. Here a different design with just the standard DCE consisting of only WAItE states was generated and two forms of external anchoring were considered. Although the model estimates for this design performed the best in terms of the number of WAItE parameter estimates that were negative, this was the only design that was generated for this type of model. Nevertheless, the results of this model were encouraging. The two anchoring strategies using VAS and pairwise anchoring also generated encouraging results. The VAS value for the PITS state was a lot higher than the value for the PITS state generated from the pairwise anchoring, this may have been because there was no trading opportunity in the VAS task.

Existing studies employing the DCE_{TTO} approach have utilised conditional logit or probit models (Bansback et al., 2012 and Norman et al., 2012). Like the multinominal logit model, the advantage of using these models is that they account for the panel nature of data (i.e. the estimation method allows for the fact that multiple observations come from one respondent, as allowing them to be independent would be incorrect). Although this is not the first study to utilise the multinominal logit model in the estimation of utility valued for health states (Hakim & Pathak 1999), perhaps another avenue of assessment for the current study could be the application of conditional logit or probit models and compare the results from the different models.

Conclusion

This study has shown that DCE_{TTO} could potentially be used for the elicitation of preference values for the new WAItE measure, but only with data obtained from 35-54 year olds. The other option could be to utilise the $DCE_{NoDuration}$ method that could be applied to all age groups. Further work needs to be undertaken before the main valuation study can be carried out.

REFERENCES

BAKKER, C. & VAN-DER-LINDEN, S. 1995. Health related utility measurement: an introduction. *Journal of Rheumatology,* 22**,** 1197-1199.

BANSBACK, N., BRAZIER, J., TSUCHIYA, A. & ANIS, A. 2012. Using a discrete choice experiment to estimate health state utility values. *Health Econ,* 31**,** 306-18.

BLIEMER, M. C. J., ROSE, J. M. & HESS, S. 2008. Approximation of Bayesian Efficiency in Experimental Choice Designs. *Journal of Choice Modelling,* 1**,** 98-127.

BRAZIER, J., KOLOTKIN, R., CROSBY, R. & WILLIAMS, G. 2004. Estimating a preference-based single index for the Impact of Weight on Quality of Life-Lite (IWQOL-Lite) instrument from the SF-6D. *Value in Health,* 7**,** 490-498.

BRAZIER, J., RATCLIFFE, J., SALOMON, J. & TSUCHIYA, A. 2007. *Measuring and Valuing Health Benefits for Economic Evaluation*, Oxford University Press.

CHOICEMETRICS 2012. Ngene. 1.1.1 ed.: http://www.choice-metrics.com/index.html.

COAST, J., FLYNN, T., NATARAJAN, L., SPROSTON, K., LEWIS, J., LOUVIERE, J. & PETERS, T. 2008. Valuing the ICECAP capability index for older people. *Social Science & Medicine* 67**,** 874-82.

DE-BEKKER-GROB, E., RYAN, M. & GERARD, K. 2012. Discrete choice experiments in health economics: a review of the literature. *Health Economics,* 21**,** 145-172.

DRUMMOND, M., STODDART, G. & TORRANCE, G. 1996. *Methods for the Economic Evaluation of Health Care Programs,* Oxford, United Kingdom, Oxford University Press.

FLYNN, T. N., LOUVIERE, J. J., MARLEY, A. A., COAST, J. & PETERS, T. J. 2008. Rescaling quality of life values from discrete choice experiments for use as QALYs: a cautionary tale. *Population Health Metrics,* 6.

HAKIM, Z. & PATHAK, D. S. 1999. Modelling the EuroQol data: a comparison of discrete choice conjoint and conditional preference modelling. *Health Econ,* 8**,** 103-116.

HEALTH-SURVEY-FOR-ENGLAND. 2008. *Table 3 - Body mass index (BMI), by survey year, age and sex* [Online]. Available: http://www.ic.nhs.uk/webfiles/publications/HSE07/ADULT%20TREND%20TABLES%202007.xls [Accessed 2010].

KOLOTKIN, R., ZELLER, M., MODI, A., SAMSA, G., QUINLAN, N., YANOVSKI, J. & AL., E. 2006. Assessing weight-related quality of life in adolescents. *Obesity,* 14**,** 448-457.

LANCASTER, K. 1966. A New Approach to Consumer Theory. Journal of Political Economy. *Journal of Political Economy,* 74**,** 132-157.

LANCSAR, E., LOUVIERE, J. & 2008; 2008. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmacoeconomics,* 26**,** 661-77.

MCFADDEN, D. 1974. Conditional logit analysis of qualitative choice behavior. *In:* ZAREMBKA, P. (ed.) *Frontiers in Econometrics.* New York: Academic Press.

NATIONAL-INSTITUTE-FOR-CLINICAL-EXCELLENCE-(NICE) December 2006. Obesity guidance on the prevention, identification, assessment and management of overweight and obesity in adults and children.

NATIONAL-INSTITUTE-FOR-CLINICAL-EXCELLENCE-(NICE) June 2008. Guide to the Methods of Technology Appraisal.

NORMAN, R., CRONIN, P. A. & VINEY, R. C. 2012. Deriving utility weights for the EQ-5D-5L using a discrete choice experiment. *CHERE Working Papers.*

ORME, B. 1998. Sample Size Issues for Conjoint Analysis Studies. *Sawtooth Software Technical Papers* [Online].

RYAN, M., BATE, A., EASTMOND, C. J. & LUDBROOK, A. 2001a. Use of discrete choice experiments to elicit preferences. *Qual Health Care,* Suppl 1.

RYAN, M., SCOTT, D., REEVES, C., BATE, A., TEIJLINGEN, E. V., RUSSELL, E., NAPPER, M. & ROBB, C. 2001b. Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technology Assessment,* 5**,** 186.

RYAN, M. & GERARD, K. 2003. Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Appl Health Econ Health Policy,* 2**,** 55-64.

RYAN, M., NETTEN, A., SKATUN, D. & 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**,** 927-944.

THE-HEALTH-AND-SOCIAL-CARE-INFORMATION-CENTRE December 2009. National Child Measurement Programme: England, 2008/09 school year. Department-of-Health.

VINEY, R., SAVAGE, E., KING, M. & HOSSAIN, I. 2007. Using Choice Experiments to Estimate QALYs: An Application to Prostate Cancer.

iHEA 2007 6th World Congress: Explorations in Health Economics Paper
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=994063.

ZELLER, M. & MODI, A. 2009. Development and initial validation of an obesity-specific quality-of-life measure for children: sizing me up. *Obesity,* Jun 17**,** 1171-7.

Table 1: Attributes and levels used to design the DCE study

Summary of attributes			Levels				
			Never	Almost never	Some-times	Often	Always
QoL	1. Tired		1	2	3	4	5
	2. Walking with others		1	2	3	4	5
	3. Avoid playing sports		1	2	3	4	5
	4. Concentrate on studies / work		1	2	3	4	5
	5. Treated differently		1	2	3	4	5
	6. Uncomfortable shopping		1	2	3	4	5
	7. Unable to do the same things as others		1	2	3	4	5
Duration	8. Life expectancy (3 different variants)	D10	10yrs	9yrs	8yrs	6yrs	5yrs
		D20	20 yrs	18 yrs	16 yrs	12 yrs	10 yrs
		D50	50 yrs	45 yrs	40 yrs	30 yrs	25 yrs

Table 2: Description of survey responses

	Pilot A		Pilot B		Pilot C	
	N	**%**	**N**	**%**	**N**	**%**
Eligible starts	128	100	277	100	352	100
Screen-outs (completed the survey within the minimum time limit)	8	6	33	12	52	14.8
Dropouts (Outside of DCE exercise)	5	4	31	11	119	33.8
Dropouts (During DCE exercise)	3	2	4	1	10	2.84
Total number who completed the full survey but failed to formally sign out from the survey and to be counted	0	0	3	1	0	0
Total number excluded from the final sample	16	13	71	26	181	51.4
Final sample available for analysis	112	88	206	74	171	48.6

Table 3: Characteristics of respondents in the surveys

Characteristic		Pilot A (N=112)		Pilot B (N=206)		Pilot C (N=171)	
		N	%	N	%	N	%
Block/version	1	15	13.4	19	9.2	17	9.9
	2	16	14.3	19	9.2	18	10.5
	3	12	10.7	29	14.1	19	11.1
	4	10	8.9	30	14.6	20	11.7
	5	12	10.7	20	9.7	19	11.1
	6	13	11.6	27	13.1	19	11.1
	7	11	9.8	19	9.2	22	12.9
	8	13	11.6	24	11.7	18	10.5
	9	10	8.9	19	9.2	19	11.1
Gender	Male	57	50.9	102	49.5	85	49.7
	Female	55	49.1	104	50.5	86	50.3
Age group	18-24	16	14.3	24	11.7	26	15.2
	25-34	35	31.3	38	18.4	31	18.1
	35-44	14	12.5	48	23.3	41	24.0
	45-54	23	20.5	55	26.7	40	23.4
	55-64	16	14.3	29	14.1	20	11.7
	65-90	8	7.1	12	5.8	13	7.6
Marital status	Married/partner	69	61.6	115	55.8	102	59.6
	Other	43	38.4	91	44.2	69	40.4
Employment status:							
Employed/self employed		64	57.1	135	65.5	114	66.7
Student		8	7.1	8	3.9	8	4.7
Not working		40	35.7	63	30.6	49	28.7
Highest level of education:							
No qualifications		5	4.5	9	4.4	8	4.7
University undergraduate / Higher degree		37	33.0	86	41.7	59	34.5
Other		70	62.5	111	53.9	104	60.8
Self assessed health	Excellent	13	11.6	28	13.6	25	14.6
	Good	59	52.7	103	50.0	84	49.1
	Fair	32	28.6	52	25.2	45	26.3
	Poor	8	7.1	19	9.2	15	8.8
	Very Poor	0	.0	4	1.9	2	1.2
Weight status*	Under wt	6	5.4	9	4.4	6	3.5
	Normal wt	41	36.6	83	40.3	69	40.4
	Overweight	41	36.6	67	32.5	50	29.2
	Obese	24	21.4	47	22.8	46	26.9
Difficulty with DCE task:							
Very difficult		12	11	25	12	23	13.5
Quite difficult		43	38	98	48	95	55.6
Nether difficult or easy		31	28	58	28	37	21.6
Fairly easy		26	23	25	12	16	9.4
Understanding DCE task:							
Fully understand		84	75	153	74	116	67.8
Partially understand		26	23	49	24	54	31.6
Did not understand		2	2	4	2	1	.6
WAItE score	Mean	17.24		16.89		17.45	
	Median	17.00		16.00		17.00	
	Std. Deviation	5.2		5.2		5.6	
	Minimum	8		7		7	
	Maximum	30		34		34	

*1=Under (BMI 18.5 or less), 2=Normal wt (BMI 18.6 to 24.9, 3=Overweight (BMI 25 to 24.9) & 4=Obese (BMI 30 plus)

Table 4: Parameter estimates - Main effects models*

Variable	Pilot A		Pilot B - Sample 1		Pilot B - Sample 2		Pilot B - Sample 3		Pilot B_ Samples 1 to 3	
	18 + yrs (D10)[a]	P-value	18-54yrs (D50)[a]	p-value	54+yrs (D20)[a]	p-value	35-54yrs (D20)[a]	p-value	18+yrs (D50) (D20)[a]	P-value
W_1_2	0.232	(0.24)	0.244	(0.23)	0.644	(0.12)	0.0419	(0.9)	**0.284**	**(0.06)**
W_1_3	0.0818	(0.79)	0.0609	(0.84)	0.666	(0.35)	**-1.05**	**(0.05)**	-0.0072	(0.98)
W_1_4	0.146	(0.74)	-0.129	(0.76)	0.82	(0.36)	-0.911	(0.19)	-0.0633	(0.83)
W_1_5	-0.102	(0.88)	0.0664	(0.91)	0.717	(0.62)	**-2.15**	**(0.05)**	-0.0603	(0.9)
W_2_2	0.321	(0.1)	0.3	(0.11)	-0.146	(0.66)	**-0.541**	**(0.03**	0.129	(0.34)
W_2_3	0.355	(0.26)	0.294	(0.32)	0.663	(0.31)	**-1**	**(0.06)**	0.161	(0.46)
W_2_4	0.433	(0.34)	0.257	(0.53)	0.0974	(0.92)	**-1.58**	**(0.05)**	-0.0354	(0.91)
W_2_5	0.543	(0.4)	0.399	(0.5)	0.578	(0.7)	**-1.94**	**(0.07)**	0.169	(0.72)
W_3_2	0.225	(0.23)	0.233	(0.2)	0.243	(0.5)	0.22	(0.45)	0.181	(0.18
W_3_3	0.445	(0.18)	0.272	(0.35)	-0.147	(0.85)	-0.263	(0.6)	0.153	(0.5)
W_3_4	0.384	(0.39)	0.281	(0.5)	0.709	(0.49)	-0.753	(0.28)	0.159	(0.61)
W_3_5	0.331	(0.62)	0.385	(0.53)	0.739	(0.64)	-1.39	(0.2)	0.2	(0.67
W_4_2	0.201	(0.31)	0.0294	(0.88)	0.0991	(0.8)	**-0.991**	**(0)**	-0.13	(0.36)
W_4_3	0.184	(0.56)	0.012	(0.97)	0.632	(0.29)	**-0.851**	**(0.1)**	-0.0155	(0.94)
W_4_4	0.35	(0.42)	0.0951	(0.81)	-0.0974	(0.91)	**-2.1**	**(0)**	-0.309	(0.3)
W_4_5	0.215	(0.75)	0.0292	(0.96)	0.417	(0.77)	**-2.66**	**(0.01)**	-0.307	(0.51)
W_5_2	0.0799	(0.71)	0.211	(0.26)	0.445	(0.44)	-0.485	(0.18)	0.114	(0.46)
W_5_3	0.0295	(0.93)	0.0967	(0.72)	0.694	(0.39)	**-1.21**	**(0.02)**	0.0509	(0.82)
W_5_4	0.107	(0.83)	0.251	(0.53)	0.472	(0.67)	**-1.74**	**(0.02)**	-0.0146	(0.96)
W_5_5	-0.0769	(0.91)	0.213	(0.73)	0.532	(0.75)	**-2.21**	**(0.04)**	-0.0613	(0.9)
W_6_2	0.419	(0.02)	**0.328**	**(0.09)**	-0.0448	(0.91)	-0.0774	(0.82)	0.208	(0.16)
W_6_3	0.203	(0.51)	0.0125	(0.96)	0.67	(0.4)	**-1.3**	**(0.01)**	-0.0919	(0.67)
W_6_4	0.044	(0.92)	-0.14	(0.73)	0.939	(0.37)	**-1.76**	**(0.02)**	-0.204	(0.52)
W_6_5	-0.069	(0.92)	-0.106	(0.86)	0.516	(0.74)	**-2.92**	**(0.01)**	-0.335	(0.48)
W_7_2	0.0989	(0.65)	-0.0129	(0.95)	0.467	(0.3)	**-0.847**	**(0.01)**	-0.0381	(0.8)
W_7_3	0.148	(0.65)	0.194	(0.5)	1.14	(0.16)	**-1.17**	**(0.03)**	0.065	(0.78)
W_7_4	0.0946	(0.83)	0.0816	(0.83)	0.475	(0.68)	**-1.6**	**(0.02)**	-0.122	(0.69)
W_7_5	-0.0237	(0.97)	-0.0582	(0.92)	0.56	(0.72)	**-2.19**	**(0.05)**	-0.293	(0.53)
D	0.197	(0.26)	0.0342	(0.29)	0.116	(0.58)	**0.472**	**(0)**	-	-
D2	-	-	-	-	-	-	-	-	**0.449**	**(0.01)**
D3	-	-	-	-	-	-	-	-	**0.861**	**(0.03)**
D4	-	-	-	-	-	-	-	-	**1.08**	**(0.04)**
D5	-	-	-	-	-	-	-	-	**1.52**	**(0.01)**
asc1	-0.00658	(0.91)	-0.0421	(0.5)	-0.176	(0.2)	**0.253**	**(0.05)**	-0.0287	(0.56)

Main effects models	Pilot A	Pilot B Sample 1	Pilot B Sample 2	Pilot B Sample 3	Pilot B Sample 1 to 3
	18 + yrs (D10)[a]	18-54yrs (D50)[a]	54+yrs (D20)[a]	35-54yrs (D20)[a]	18+yrs (D50)[a] (D20)[a]
No. observations	1100	1130	410	480	2020
No. individuals	110	113	41	48	202
Likelihood ratio test	125.896	99.9	123.29	102.412	228.549
Adjusted rho-square	0.043	0.025	0.111	0.064	0.058
No.-ve QoL coefficients (%)	4 (14)	5 (18)	4 (14)	26 (93)	16 (57)

*The analysis excludes all those who self report to not understanding the DCE tasks. P<0.10 for coefficients in **bold**. [a] D10 = standard duration levels, D20 = standard duration levels scaled up by 2 and D50 = standard duration levels scaled up by 5

Table 5: Choice models - Interaction effects*

Variable	Pilot A		Pilot B - Sample 1		Pilot B - Sample 2		Pilot B - Sample 3		Pilot B - Samples 1 to3	
	18 + yrs (D10)[a]	**P-value**	**18-54yrs (D50)[a]**	**p-value**	**54+yrs (D20)[a]**	**p-value**	**35-54yrs (D20)[a]**	**p-value**	**18+yrs (D50)[a] (D20)[a]**	**P-value**
W1_2_D	0.0359	0.13	**0.0085**	**0.06**	**0.0767**	**0**	0.0264	0.18	**0.00714**	**0.08**
W1_3_D	0.0187	0.52	0.00266	0.64	**0.0961**	**0**	-0.0224	0.29	-0.00268	0.52
W1_4_D	0.0271	0.51	0.0003	0.97	**0.123**	**0**	0.0118	0.74	-0.00653	0.18
W1_5_D	0.00179	0.98	0.00614	0.58	**0.158**	**0**	-0.0404	0.45	-0.00673	0.21
W2_2_D	**0.0444**	**0.04**	**0.0082**	**0.04**	0.0318	0.12	-0.019	0.3	0.00439	0.19
W2_3_D	**0.0523**	**0.06**	0.00866	0.1	**0.0971**	**0**	-0.0211	0.48	0.00284	0.46
W2_4_D	**0.0666**	**0.09**	0.00747	0.35	**0.0715**	**0.04**	-0.0354	0.3	-0.00332	0.43
W2_5_D	0.0823	0.14	0.0128	0.22	**0.138**	**0**	-0.035	0.5	0.00131	0.8
W3_2_D	0.0304	0.16	0.0059	0.16	**0.0383**	**0.03**	0.034	0.12	0.00338	0.33
W3_3_D	**0.0585**	**0.03**	0.00711	0.18	**0.0427**	**0.2**	0.029	0.26	0.00191	0.58
W3_4_D	0.0536	0.16	0.00937	0.17	**0.116**	**0**	0.0193	0.57	0.00176	0.68
W3_5_D	0.0474	0.41	0.0122	0.27	**0.152**	**0.01**	0.00393	0.94	5.22E-05	0.99
W4_2_D	0.0257	0.24	0.000982	0.82	**0.0357**	**0.08**	**-0.0521**	**0.01**	-0.00343	0.36
W4_3_D	0.0272	0.4	0.00301	0.63	**0.0943**	**0**	-0.0241	0.43	-0.00237	0.6
W4_4_D	0.0519	0.19	0.00536	0.49	**0.0647**	**0.1**	**-0.0862**	**0.01**	**-0.00808**	**0.09**
W4_5_D	0.039	0.46	0.00413	0.68	**0.121**	**0.01**	-0.0814	0.1	**-0.0116**	**0.03**
W5_2_D	0.0137	0.53	0.00593	0.14	0.0293	0.27	-0.00166	0.93	0.00226	0.54
W5_3_D	0.00495	0.87	0.00283	0.6	**0.0677**	**0.02**	-0.0376	0.15	-0.00134	0.74
W5_4_D	0.0147	0.72	0.00814	0.27	**0.0818**	**0.06**	-0.0507	0.11	-0.000528	0.92
W5_5_D	-0.00705	0.9	0.00857	0.39	**0.131**	**0.04**	-0.0399	0.44	-0.00532	0.27
W6_2_D	**0.0514**	**0.01**	**0.00963**	**0.01**	**0.0251**	**0.15**	0.0278	0.12	0.00568	0.15
W6_3_D	0.03	0.26	0.00119	0.81	**0.0897**	**0**	-0.0357	0.18	**-0.00615**	**0.08**
W6_4_D	0.015	0.72	-0.00086	0.9	**0.127**	**0**	-0.0521	0.18	**-0.0112**	**0.01**
W6_5_D	-0.00474	0.93	0.000779	0.94	**0.142**	**0.01**	-0.0832	0.11	**-0.0136**	**0.01**
W7_2_D	0.0205	0.33	0.000947	0.8	**0.0488**	**0.01**	-0.024	0.1	-0.00235	0.5
W7_3_D	0.0212	0.47	0.0079	0.16	**0.123**	**0**	-0.0234	0.33	0.00131	0.73
W7_4_D	0.0164	0.67	0.00488	0.49	**0.111**	**0.01**	-0.0489	0.1	-0.00487	0.2
W7_5_D	0.016	0.78	0.00353	0.74	**0.15**	**0.02**	-0.0464	0.38	**-0.0122**	**0.02**
D	0.00284	0.99	-0.00321	0.95	**-0.615**	**0.03**	0.423	0.1	-	-
D2	-	-	-	-	-	-	-	-	**0.511**	**0**
D3	-	-	-	-	-	-	-	-	**1.04**	**0**
D4	-	-	-	-	-	-	-	-	**1.37**	**0**
D5	-	-	-	-	-	-	-	-	**1.93**	**0**
asc1	0.0195	0.75	-0.0294	0.66	-0.104	0.47	0.178	0.11	-0.0311	0.53

Interaction effects models	Pilot A	Pilot B Sample 1	Pilot B Sample 2	Pilot B Sample 3	Pilot B Sample 1 to 3
	18 + yrs (D10)[a]	18-54yrs (D50)[a]	54+yrs (D20)[a]	35-54yrs (D20)[a]	18+yrs (D50)[a] (D20)[a]
No. observations	1100	1130	410	480	2020
No. individuals	110	113	41	48	202
Likelihood ratio test	127.757	103.547	127.233	99.534	224.071
Adjusted rho-square	0.044	0.028	0.118	0.059	0.056
No.-ve QoL coefficients (%)	2 (7)	**NA**	**NA**	20 (71)	17 (61)

*The analysis excludes all those who self report to not understanding the DCE tasks. P<0.10 for coefficients in **bold**. [a] D10 = standard duration levels, D20 = standard duration levels scaled up by 2 and D50 = standard duration levels scaled up by 5

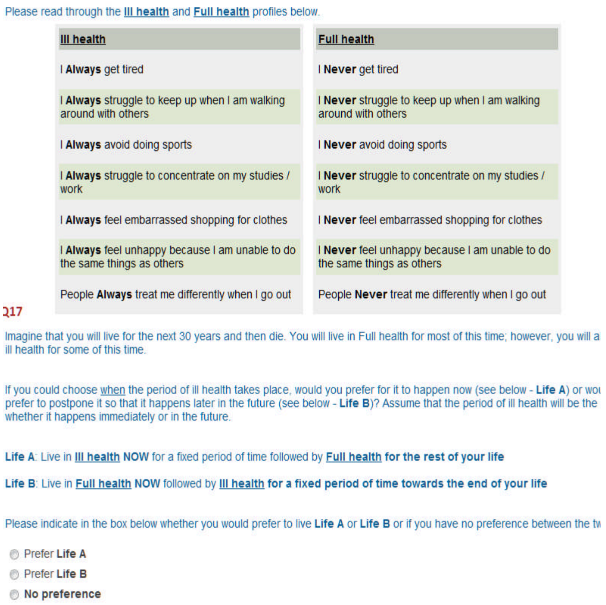
Table 6: Parameter estimates - Main effects models No Duration attribute*

Variable	Pilot A		Pilot B - Sample 1		Pilot B - Sample 2		Pilot B - Sample 3		Pilot B - Sample 1 to 3		Pilot C	
	18 + yrs (D10)[a]	**P-value**	**18-54yrs (D50)[a]**	**p-value**	**54+yrs (D20)[a]**	**p-value**	**35-54yrs (D20)[a]**	**p-value**	**18+yrs (D50)[a] (D20)[a]**	**P-value**	**18 + yrs (No D)[a]**	**P-value**
W_1_2	0.408	0.53	0.617	0.49	2.5	1	-9.49	1	0.175	0.58	0.0216	0.88
W_1_3	0.597	0.85	-0.259	0.7	4.37	1	-7.77	1	-0.0659	0.92	-0.112	0.62
W_1_4	0.286	0.84	**-1.01**	**0.01**	-2.29	1	0.285	1	-0.153	0.91	-0.392	0.23
W_1_5	0.22	0.69	-0.314	0.84	-0.194	1	-3.31	1	-0.0137	1	-0.689	0.11
W_2_2	-0.103	0.97	0.61	0.49	5.26	1	3.06	1	-0.0897	0.89	0.0656	0.59
W_2_3	0.0488	0.97	0.0972	0.94	-3.11	1	1.2	1	0.179	0.8	-0.254	0.26
W_2_4	0.446	0.83	-0.736	0.19	-1.4	1	-3.77	1	-0.269	0.79	-0.21	0.52
W_2_5	0.166	0.94	-0.504	0.71	5.67	1	-1.19	1	-0.178	0.27	-0.49	0.25
W_3_2	-0.47	0.77	0.719	0.71	5.01	1	1.61	1	0.108	0.68	0.147	0.29
W_3_3	-0.461	0.87	0.79	1	0.559	1	2.44	1	0.191	0.64	-0.0351	0.87
W_3_4	-0.126	0.9	0.896	1	0.244	1	-6.33	1	0.242	0.84	-0.0412	0.9
W_3_5	-0.16	0.95	0.463	0.77	1.07	1	-2.38	1	0.226	0.57	-0.286	0.51
W_4_2	-0.228	0.92	0.248	0.88	-0.858	1	-2.11	1	0.0547	0.93	-0.0734	0.6
W_4_3	0.2	0.93	-0.224	0.89	1.91	1	4.33	1	0.142	0.83	-0.0229	0.92
W_4_4	0.312	0.92	-0.0826	0.97	-3.29	1	-2.47	1	-0.181	0.01	-0.22	0.51
W_4_5	0.437	0.8	-0.575	0.4	0.408	1	4.39	1	-0.154	0.73	-0.563	0.21
W_5_2	0.662	0.58	-0.32	0.5	1.43	1	-3.33	1	0.0432	0.97	0.13	0.32
W_5_3	-0.3	0.91	0.167	0.94	3.31	1	6.17	1	0.0842	0.92	-0.237	0.3
W_5_4	-0.269	0.82	0.261	0.8	2.83	1	-6.14	1	-0.0158	0.98	-0.184	0.57
W_5_5	-0.511	0.48	0.885	0.6	-4.03	1	-1.7	1	0.117	0.72	-0.36	0.4
W_6_2	0.64	0.62	0.0594	0.99	-8.04	1	-1.75	1	0.149	0.91	0.00234	0.99
W_6_3	0.0309	0.99	0.149	0.67	6.11	1	-1.01	1	-0.0457	0.94	-0.09	0.7
W_6_4	-0.375	0.67	-0.15	0.9	10	1	-1.64	1	-0.179	0.9	-0.282	0.39
W_6_5	-0.0791	0.93	0.203	0.71	-5.01	1	0.94	1	-0.11	0.94	-0.637	0.15
W_7_2	0.703	0.27	0.145	0.89	-0.691	1	-2.09	1	0.204	0.92	-0.0991	0.49
W_7_3	0.185	0.91	-0.187	0.91	4.16	1	-0.462	1	-0.0127	0.99	**-0.382**	**0.07**
W_7_4	-0.188	0.95	-0.172	0.83	-2.5	1	8.82	1	-0.156	0.81	-0.51	0.12
W_7_5	-0.569	0.77	0.119	0.69	-6.82	1	10	0	0.109	0.89	**-0.864**	**0.04**
asc1	-0.0393	0.97	0.0282	0.98	-3.05	1	4.61	1	0.0438	0.94	-0.0889	0.13

Main effects models	**Pilot A**	**Pilot B Sample 1**	**Pilot B Sample 2**	**Pilot B Sample 3**	**Pilot B Sample 1 to 3**	**Pilot C**
	18 + yrs (D10)[a]	18-54yrs (D50)[a]	54+yrs (D20)[a]	35-54yrs (D20)[a]	18+yrs (D50)[a] (D20)[a]	18+yrs (No D)
No. observations	237	239	106	112	457	1700
No. individuals	110	113	41	48	202	170
Likelihood ratio test	32.911	32.236	26.278	33.992	21.306	90.905
Adjusted rho-square	-0.076	-0.078	-0.216	-0.155	-0.058	0.014
No.-ve QoL coefficients (%)	13 (46)	12 (43)	12 (43)	17 (61)	14 (50)	23 (82)

*The analysis excludes all those who self report to not understanding the DCE tasks. P<0.10 for coefficients in **bold**. [a] D10 = standard duration levels, D20 = standard duration levels scaled up by 2 and D50 = standard duration levels scaled up by 5

Figure 1: Example of time preference question



Please read through the **Ill health** and **Full health** profiles below.

Ill health	Full health
I **Always** get tired	I **Never** get tired
I **Always** struggle to keep up when I am walking around with others	I **Never** struggle to keep up when I am walking around with others
I **Always** avoid doing sports	I **Never** avoid doing sports
I **Always** struggle to concentrate on my studies / work	I **Never** struggle to concentrate on my studies / work
I **Always** feel embarrassed shopping for clothes	I **Never** feel embarrassed shopping for clothes
I **Always** feel unhappy because I am unable to do the same things as others	I **Never** feel unhappy because I am unable to do the same things as others
People **Always** treat me differently when I go out	People **Never** treat me differently when I go out

Q17

Imagine that you will live for the next 30 years and then die. You will live in Full health for most of this time; however, you will a ill health for some of this time.

If you could choose when the period of ill health takes place, would you prefer for it to happen now (see below - **Life A**) or wou prefer to postpone it so that it happens later in the future (see below - **Life B**)? Assume that the period of ill health will be the whether it happens immediately or in the future.

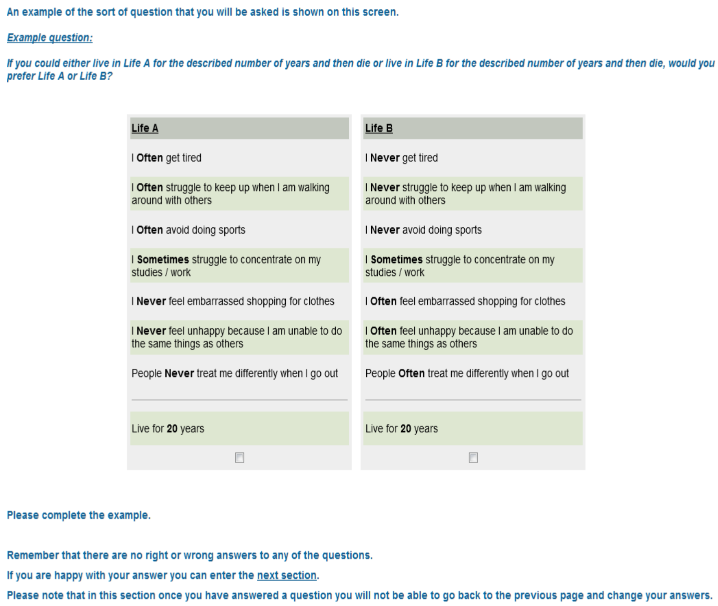
Life A: Live in **Ill health** NOW for a fixed period of time followed by **Full health for the rest of your life**

Life B: Live in **Full health** NOW followed by **Ill health for a fixed period of time towards the end of your life**

Please indicate in the box below whether you would prefer to live **Life A** or **Life B** or if you have no preference between the tw

- Prefer **Life A**
- Prefer **Life B**
- **No preference**

Figure 2: Example of DCE question (with duration)*



An example of the sort of question that you will be asked is shown on this screen.

Example question:

If you could 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
I **Often** get tired	I **Never** get tired
I **Often** struggle to keep up when I am walking around with others	I **Never** struggle to keep up when I am walking around with others
I **Often** avoid doing sports	I **Never** avoid doing sports
I **Sometimes** struggle to concentrate on my studies / work	I **Sometimes** struggle to concentrate on my studies / work
I **Never** feel embarrassed shopping for clothes	I **Often** feel embarrassed shopping for clothes
I **Never** feel unhappy because I am unable to do the same things as others	I **Often** feel unhappy because I am unable to do the same things as others
People **Never** treat me differently when I go out	People **Often** treat me differently when I go out
Live for **20** years	Live for **20** years

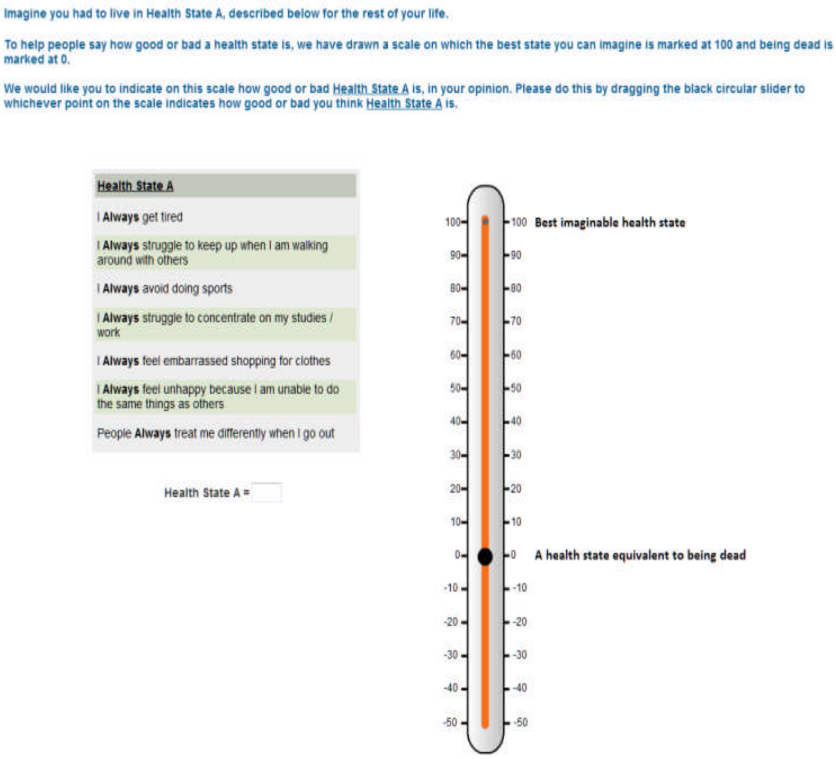
Please complete the example.

Remember that there are no right or wrong answers to any of the questions.

If you are happy with your answer you can enter the next section.

Please note that in this section once you have answered a question you will not be able to go back to the previous page and change your answers.

Figure 3: Example of VAS question



Imagine you had to live in Health State A, described below for the rest of your life.

To help people say how good or bad a health state is, we have drawn a scale on which the best state you can imagine is marked at 100 and being dead is marked at 0.

We would like you to indicate on this scale how good or bad Health State A is, in your opinion. Please do this by dragging the black circular slider to whichever point on the scale indicates how good or bad you think Health State A is.

Health State A
I **Always** get tired
I **Always** struggle to keep up when I am walking around with others
I **Always** avoid doing sports
I **Always** struggle to concentrate on my studies / work
I **Always** feel embarrassed shopping for clothes
I **Always** feel unhappy because I am unable to do the same things as others
People **Always** treat me differently when I go out

Health State A =

Figure 4: Valuation of the PITS state using VAS Anchoring

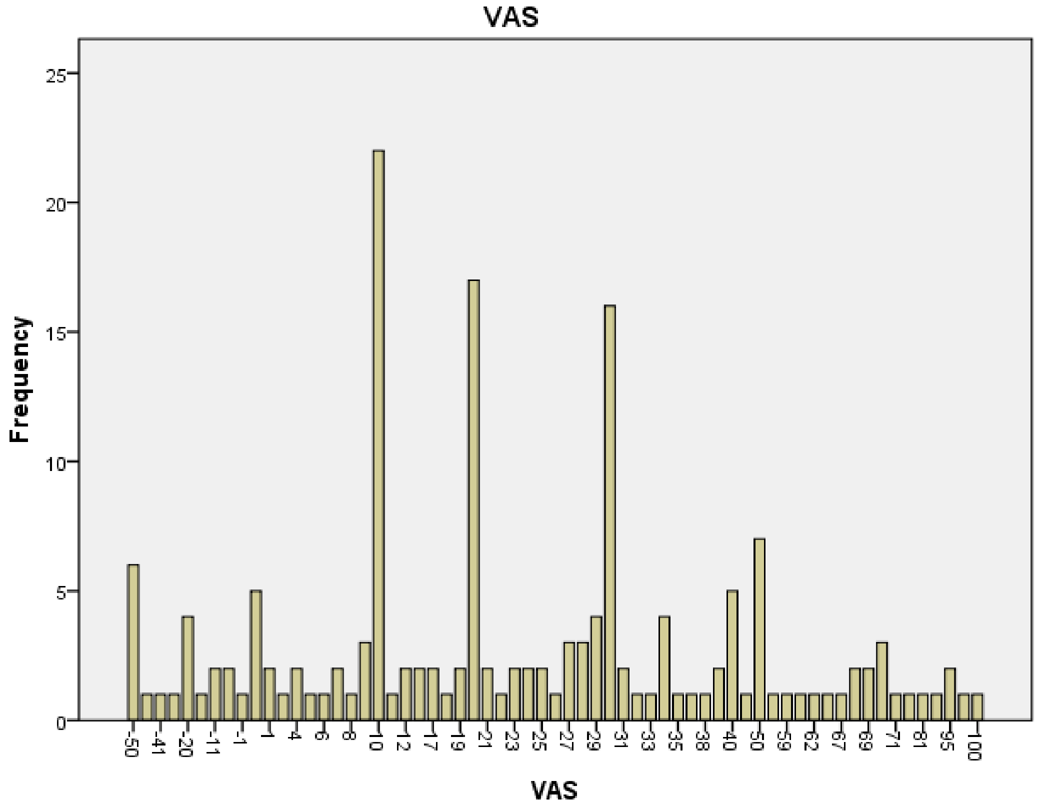


Figure 5: - Valuation of the PITS state using pairwise Anchoring