Exploring differences between TTO and DCE in the valuation of health states

Robinson, Angela; Spencer, Anne E.; Pinto Prades, Jose Luis; Covey, Judith A.

Published in:
Medical Decision Making

DOI:
10.1177/0272989X16668343

Publication date:
2017

Document Version
Author accepted manuscript

Citation for published version (Harvard):
Robinson, A, Spencer, AE, Pinto Prades, JL & Covey, JA 2017, 'Exploring differences between TTO and DCE in the valuation of health states', *Medical Decision Making*, vol. 37, no. 3, pp. 273-284. https://doi.org/10.1177/0272989X16668343

Exploring differences between TTO and DCE in the valuation of health states.

Angela Robinson[a], Anne Spencer[b], Jose-Luis Pinto-Prades[c], Judith Covey[d]

a. Norwich Medical School,
University of East Anglia,
Earlham Road,
Norwich,
NR4 7TJ
angela.robinson@uea.ac.uk (corresponding author)

b. Exeter Medical School,
University of Exeter,
Exeter,
EX2 4SG
a.e.spencer@exeter.ac.uk

c. Department of Economics,
Universidad de Navarra
C/ PIO XII, 36, Pamplona 31008
and
Yunus Centre for Social Business & Health,
Glasgow Caledonian University,
58 Port Dundas Road.
Glasgow
G4 0BA
joseluis.pintoprades@gmail.com

d. Dept of Psychology
Durham University
Queen's campus,
Stockton on Tees
TS176BH
j.a.covey@durham.ac.uk

Financial support for this study was provided by a grant to Jose Luis Pinto Prades from the Spanish Ministry of Education, Research Grant ECO2010.22041.C02.01 and Junta de Andalucía, Project P09.SEJ.4992. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

ABSTRACT

There is recent interest in using Discrete Choice Experiments (DCEs) to derive health state utility values and results can differ from Time Trade Off (TTO). Clearly DCE is 'choice-based' whereas TTO is generally considered to be a 'matching' task. We explore whether procedural adaptations to the TTO -which make the method more closely resemble a DCE -makes TTO and choice converge. In particular, we test whether making the matching procedure in TTO less 'transparent' to the respondent reduces disparities between TTO and DCE. We designed an interactive survey that was hosted on the internet and 2022 interviews were achieved in the UK in a representative sample of the population. We found a marked divergence between TTO and DCE, but this was not related to the 'transparency' of the TTO procedure. We conclude that a difference in the error structure between TTO and choice and that factors other than differences in utility are affecting choices is driving the divergence. The latter has fundamental implications for the way choice data are analysed and interpreted.

Keywords: Utility assessment, Matching, Choice, EQ-5D 5L, DCE, TTO,

INTRODUCTION

There has been recent interest in the use of discrete choice experiments (DCE) to derive health state utilities for use in QALY calculations[1-5]. When compared head to head, DCE and TTO have been shown to arrive at different utility estimates[3,6] but to date little research has gone into exploring the factors that might be driving differences. Arguments have been put forward previously regarding the relative merits of DCE compared with 'traditional' methods such as TTO and SG. For example, it has been argued that traditional value elicitation techniques, such as TTO and SG, that set out to establish an individual's point of indifference are more cognitively demanding than those involving pair-wise choices[4,5]. For example, in discussing the TTO, Bansback and colleagues[1] consider "there is still a concern that the tasks involved are still too cognitively demanding for certain populations, resulting in response inconsistencies and subsequent data exclusions, which limit the representativeness of the values obtained" (p. 306).

There are, of course, a number of features of an actual DCE that may explain differences in valuations across methods, such as the functional form of the model deployed in modelling the choice data. We are interested here, however, in the choices themselves and not how the choice data are subsequently modelled. There are many reasons why a DCE that sets out to value health state utilities using time as the numeraire –sometimes referred to as DCE_{TTO} - may yield systematically different valuations from traditional TTO.

It has been observed for some time that preferences over two options can change depending on the elicitation procedure used. In particular, it is well known that 'matching' and 'choice' tasks yield different results[7] and this has been studied previously in relation to health state utility measurement[8-10]. In one study using TTO, Attema and colleagues found significantly different valuations between matching and choice along with the presence of preference reversals[10]. Matching may encourage more quantitative decision making processes and give more weight to the attribute used as the 'currency' on which to match, whilst choice may encourage more qualitative decision making and give more weight to the most 'prominent' attribute[7,11]. Whilst, on the face of it, this may appear to offer an explanation of any differences between DCE and more traditional utility elicitation methods, in reality TTO (and SG) are generally operationalised as a series of pair-wise choices that set out to 'home in' on a point of indifference, a technique which has been referred to as 'choice-based matching'[12]. Fischer and colleagues developed the task-goal hypothesis and argued that more weight is given to the prominent attribute when the aim is to differentiate between options (as in choice), than in tasks where the aim is to equate options (as in matching). They showed that when the objective – or goal – of the matching task was made less transparent using a 'hidden choice-based matching' technique, preference reversals between matching and choice were reduced[12].

Applying this to the context here, it will generally be obvious to respondents that a TTO is iterating towards a point of indifference and they are being asked via this process to 'match' a number of years in normal health to X years in the target health state. This aim could, however, be made less transparent to respondents and the findings of Fischer lead us to hypothesise this will reduce differences between TTO and direct choice. This essentially involves using an iterative procedure to arrive at the point of indifference in TTO, but moves away from valuing states sequentially where one state is valued before moving on to the next. In contrast, states could be valued concurrently, whereby the respondent sees different states in alternating questions (this will be explained in detail below). Arguably, the task in TTO would be even less transparent if non-iterative procedures were used to arrive at the point(s) of indifference. Valuing health states concurrently in TTO using a non-iterative procedure is more in line with how states would be valued within a DCE.

The issue of interest here is whether we can predict choices between health profiles based on respondent's TTO valuations. Such a prediction clearly requires the imposition of restrictions on the utility function for health. We could assume, for example, that the linear QALY model holds and simply estimate the total number of QALYs in each alternative and predict that the respondent will choose the alternative offering the higher number. Linearity is, however, a very restrictive assumption, so we rely here on the weaker conditions of mutual utility independence (MUI) and constant proportional trade-off (CPTO) which allow subjects to discount future health. More details about these assumptions are given in Appendix 1. We then test whether choices can be predicted from TTO responses when the TTO procedure varies according to how 'transparent' the TTO task is in relation to; a) whether an iterative or non-iterative procedure is used to arrive at a point of indifference and b) whether health states are valued 'sequentially' or 'concurrently'.

If varying these factors *can* offer an explanation of differences between TTO and direct choice, then we would expect doing so to; a) systematically influence the TTO valuations themselves and b) bring about convergence between TTO and direct choice.

The objectives of the current study are therefore to:

1. Examine whether TTO responses are robust to the procedural variations listed above.
2. Examine to what extent direct choices may be predicted from TTO responses.

METHODS

Survey design

In order to explore all factors of interest, but without over burdening respondents, 8 different versions of the survey were designed and hosted on the internet. Sections 1-3 of the survey were identical for all versions and are described in Appendix 2. In section 4, respondents were randomised to one of 8 versions of the survey according to which variant of TTO – and set of health states- they would see.

Before going on to explain the TTO variants in detail, we first describe the health states used in the survey. The health states were based on the EQ-5D 5L descriptive system. Two sets of health states were constructed- 'Odd' and 'Even' -which were used in the odd and even numbered groups respectively. The health states are set out in figure 1.

Figure 1: **EQ 5D (5L) health states used in the TTO exercises**

Odd	Even
11121	13122
21211	13224
12212	23242
13122	23314

The health states were chosen to cover different severities whilst minimising the likelihood that any state would be rated as worse than dead by a large number of respondents (we explain this further in explaining the 'direct choice' questions). It is easy to see, however, that the Even set is generally more severe than the Odd set. One state – 13122- was common to both groups, which offers a test of the impact of ‘context’ on valuation. Each set also included one state that strictly dominated at least one other. Thus, 11121 dominates 13122 in the Odd set and 13122 dominates 13224 and 23242 in the Even set. The inclusion of strict dominance offers a straightforward test of consistency of responses.

Iterative TTO procedures

The TTO variants may be separated broadly into ‘iterative’ and ‘non-iterative’ procedures. We begin by describing the ‘iterative’ variants in detail.

Iterative, states valued sequentially’ (i.e. traditional TTO)

This variant replicates a ‘traditional’ TTO exercise although duration is not presented graphically as some researchers have done previously. Respondents are first presented with a choice between 20 years in Life A and 10 years in Life B. The scenario as presented to respondents is depicted in Figure 2 using state 21211 as an example. If the respondent preferred 10 years in Life B to 20 years in Life A, they were then presented with a choice between 8 years in Life B and 20 years in Life A. If the respondent preferred 20 years in Life A to 10 years in Life B, they were then presented with a choice between 20 years in Life A and 12 years in Life B. This iterative process continued until they ‘switched’ to preferring Life A to Life B in successive two year intervals –or vice versa-they were then asked about the year in between. For example, if they ‘switched’ from preferring Life A to Life B between 14 years and 16 years in Life B, they were then asked about 15 years in Life B. The utility value was then taken as the midpoint of the years between which they ‘switched’. So, if they preferred Life B at 16 years, but Life A at 15 years, the utility value was taken to be 15.5/20 = 0.775. Thus, the utilities were measured to the nearest 0.025. If they still preferred Life A when the number of years in Life B was 19, they were asked about 19 years and 6 months- and then about 19 years and 9 months if they continued to prefer Life A. This was done in order to introduce greater sensitivity towards the top end of the utility space. Those respondents who would not trade even 3 months of life expectancy to avoid the health state in question were considered to be ‘non-traders’ and a value of ‘1’ was attached to that health state.

At the other end of the scale, if they still preferred one year in Life B to 20 years in Life A, they were asked whether they would prefer immediate death to 20 years in Life A. No worse than dead valuations were sought- if respondents said ‘yes’ to the ‘immediate death’ question, their valuation of that health state was taken to be zero.

Figure 2: The basic TTO scenario used (using 21211 as an example).

Please choose between the Life A and Life B shown below. Read the descriptions and numbers of lives carefully before you make a choice:

LIFE A		**LIFE B**	
20 YEARS WITH		10 YEARS WITH	
Slight problems in walking about		NO problems in walking about	
NO problems washing or dressing oneself		NO problems washing or dressing oneself	
Slight problems doing usual activities		NO problems doing usual activities	
No pain or discomfort		NO pain or discomfort	
NOT anxious or depressed		NOT anxious or depressed	
FOLLOWED BY DEATH		FOLLOWED BY DEATH	

Which would you prefer?

o Life A o Life B

Click NEXT to continue

In this iterative sequential variant the iterative procedure is followed through to the end for each health state before moving onto the next. Hence, states are valued 'sequentially' as is traditional in health state valuation exercises such as TTO and SG. The way in which this differs across the remaining iterative versions is explained below. The health states were valued in the order: 12212, 11121, 13122, 21211 in the Odd group and 23242, 13122, 23314, 13224 in the Even group.

Iterative, states valued concurrently

This is the variant that is most akin to 'hidden choice-based matching' discussed above. The main feature of this variant is that, rather than working through the iterative procedure for one health state before moving onto the next, the iterative procedures were effectively 'spliced' together and the valuations undertaken concurrently. Hence, even within an iterative procedure, it is arguably less 'transparent' to respondents what their task is for each state. For example, in the odd version respondents were first asked to consider 10 years in 12212- denoted as Life A -and 20 years in normal health-denoted as Life B. Irrespective of their response, they would next be asked to choose between a different Life A- this time 10 years in 21211 -and 20 years in Life B and so on until all 4 health states had appeared in Life A. After this first 'round' of 4 choices had been completed, the next 4 questions were each the next step in an iterative procedure underway for each health state. Each iterative procedure was identical to that described above for the traditional TTO and continued until all 4 states had been valued. The order in which the states appeared in each round was the same as they were valued in the traditional TTO.

Non-iterative versions

The other broad category of TTO variants deployed in the survey is 'non -iterative' approaches. As the name suggests, the main feature of non-iterative TTOs was that they do not set out to 'home in' on a point of indifference. Rather, respondents are presented with choices that are not based on their previous responses. The use of a non-iterative TTO procedure is arguably again making it less transparent to respondents that their task is to equate options. As with the iterative approaches, the non-iterative versions may be further classified according to whether health states are valued 'sequentially' or 'concurrently'.

Non-iterative, states valued sequentially

Respondents randomised to the 'non-iterative-sequential' variants were first asked to choose between Life A-20 years in the first health state under evaluation and Life B-either 4, 8, 12 or 16 years in normal health with that number being allocated randomly. Irrespective of their response to the first question, the number of years in Life B was changed to one of the 3 remaining durations- again drawn randomly. And so on until all 4 durations had appeared in Life B. The responses to the initial 4 questions allowed the 'range' within which that respondent's utility value lies to be estimated. The number of years in normal health in Life B was then set at the midpoint of that range. For example, consider the following sequence of responses to the first 4 questions (where the number of years in Life B relate to years in normal health).

1. 8, Life B vs 20, Life A- Prefer Life A,
2. 16, Life B vs 20, Life A- Prefer Life B,

3. 12, Life B vs 20, Life A-Prefer Life B,
4. 4, Life B vs 20, Life A- Prefer Life A.

As the respondent 'switches' from preferring Life A to preferring Life B between 8 and 12 years, they are next asked about 10 years (in normal health) in Life B. Depending on their response to that question, they would then be asked about 11 or 13 years (in normal health) in Life B. Thus, it is obvious that it is only the first part of the procedure that is truly non–iterative (n.b. a wholly non-iterative system that assessed utility values to the same degree of accuracy as in the traditional TTO, would entail presenting respondents with 20 choices for each health state). Utility values were derived in exactly the same way and recorded to the same degree of accuracy as in the traditional TTO. As in the traditional TTO, the valuation procedure is followed all the way through for each health state before moving onto the next. The order in which the states were valued in this version was the same as in the traditional TTO. Unlike with iterative procedures, respondents in the non-iterative versions *can* give inconsistent responses in a TTO for any particular health state and then no utility value may be estimated, raising the possibility of missing data.

Non-iterative, states valued concurrently.

In this variant both the number of years in Life B *and* the health state that appeared in Life A were allocated randomly. Thus, the first 16 questions respondents were a random draw from all possible combinations of durations and health states. It could be argued that it is this variant that makes the task of the TTO the least transparent and best replicates the pattern of choices that respondents would face in a DCE. Responses to these 16 questions allowed the 'range' within which that respondent's utility value for each of the 4 health states lies. The procedure thereafter was exactly as in the 'sequential' version described above and ended after all 4 health states had been valued.

Table 1 shows the 8 TTO variants and identifies which variants may be considered to be the least transparent and the most transparent.

Table 1: Summary of the TTO variants 1-8

	States valued Sequentially		**States valued Concurrently**	
	ODD	EVEN	ODD	EVEN
Iterative	**Group 1** *	**Group 2** *	**Group 3**	**Group 4**
Non-iterative	**Group 5**	**Group 6**	**Group 7** **	**Group 8** **

*The most 'transparent' TTO method (traditional TTO)
**The least 'transparent' TTO method

The direct choice questions

All respondents then answered 6 'direct choice' questions in which pairs of EQ-5D health states were compared *directly* to one another and the choice was between X years in one health state and Y years in the other. This is in contrast to the TTO whereby – irrespective of variant- the actual choice made is always between X years *in normal health* and 20 years in the 'target' health state. The relative valuation of two different 'target' health states is then inferred *indirectly* from the TTO responses. In estimating QALY gains for use in economic evaluation, however, we are generally concerned with 'moves' between one EQ 5D health state and another-so it could be argued that it is the 'direct' valuation that is the more legimate.

The basic idea behind the direct choice questions was to take an individual's TTO responses to two different EQ 5D health states and to present them with a choice between X years in one health state and Y years in the other. The values of X and Y were set such that the respondent ought to be indifferent between the two alternatives. More details of the approach are given in Appendix 2.

Suppose that U1 and U2 are the TTO utility values for health states 1 and 2 respectively. The programme would first select the state with the lower utility value. Suppose that $U1<U2$. The direct choice would present respondents with X years in health state 1 and $U1/U2 * X$ years in health state 2. In each choice, one of the two states always appeared in Life A- whilst the other appeared in Life B- and this was set in advance. Thus, either Life A or Life B could involve

the greater number of life years- depending on the respondent's valuation of the health states in the TTO. Suppose the health state that appeared in Life A had been valued more highly in the TTO than the other in the pair. Three different values of X were then used in Life B: 17, 18 and 19 years- which were assigned at random. So, for example, if U1 and U2 equalled 0.6 and 0.8 respectively, and 18 years was selected as the value of X, Y would then be set at 0.6/0.8 * 18 = 13.5 years. The respondent would then be presented with a choice between Life A: 13.5 years in health state 1 and Life B: 18 years in health state 2. Alternatively, a respondent with U1 and U2 of 0.8 and 0.6 respectively, would be faced with a choice between Life A: 18 years in health state 1 and Life B: 13.5 years in health state 2. In the case where U1=U2, then X and Y would take on the same value (either 17, 18 or 19 years) and, as above, in each pair the same health state would always appear under Life A and the other in Life B. The assumption is always that the choice has been set up such that the respondent ought to be indifferent between Lives A and B.

No direct choice question was generated whenever a respondent rated one of the health states as worse than dead or were inconsistent in the valuation of either health state such that a utility value could not be estimated (this could only happen in the non-iterative versions). We return to this issue in the results section.

ANALYSIS

We first tested whether TTO responses are robust to the procedural variations explored here. We tested procedural variations using a linear regression that allowed for the clustering of observations by respondent. The dependent variable was the TTO value. Dummies were included for the health states (6 dummies, with the base level being the common health state 13122), and TTO method (i.e. iterative concurrent, non-iterative sequential, non-iterative concurrent). The dummy variables were set up so that the constant term represented the TTO value for health state 13122 under the traditional TTO method (i.e. iterative sequential) and interaction terms were included to explore the impact of the variants on the remaining health states. We used a chi-squared test to determine whether the variant dummies and their interactions were simultaneously zero. This is similar to testing for significant differences between a model with these variant variables added and a model without them, i.e. the difference between full model and reduced models. In another version of this model, we also included a dummy for the Even - as opposed to Odd- group, to provide a means of testing whether the TTO value of the common state (13122) was subject to ‘context’ effects.

We next tested whether direct choices may be predicted from TTO responses. If responses to TTO and choice coincide we expect the ‘splits’ in the choices to be 50:50 on average and, hence the probability of choosing Life A or Life B to be 0.5. In each case, we test this by a one- sample binomial test of whether the probability of choosing Life A was significantly different than 0.5. We then pool the data across direct choice questions and test whether the pattern of choices of respondents in the other TTO variants are significantly different than those in the traditional TTO using a Generalised Estimating Equation (GEE) model that allows for clustering of observations on individuals.

RESULTS

Data were collected in June 2014 and 2022 completed interviews were achieved. The sample comprised of 947 (46.8%) males and 1075 (53.2%) females. Mean (median) age was 44.6 (45) with a range of 18-70. The age/gender breakdown is shown in Appendix 3 and compared to a representative UK population.

The mean (median) utility values derived for the health states in TTO variants 1-8 are presented in Table 2. In all variants, the general pattern of responses across *health states* is roughly as expected in that milder states are generally valued more highly than the more severe. There is no immediately obvious pattern, however, across *variants* of the TTO, but we look at the responses in more detail below.

*Table 2: Mean (median) utility values from TTO exercises by group**

TTO method	Group	11121	21211	12212	13122	13224	23242	23314
Iterative: sequential (traditional TTO)	Group 1	.813 (.925)	.787 (.925)	.652 (.775)	.667 (.775)			
	Group 2				.669 (.825)	.477 (.475)	.421 (.425)	.466 (.475)
Iterative: concurrent	Group 3	.803 (.925)	.763 (.875)	.662 (.775)	.627 (.725)			
	Group 4				.632 (.775)	.392 (.375)	.311 (.175)	.375 (.350)
Non- iterative: sequential	Group 5	.855 (.957)	.808 (.925)	.714 (.825)	.658 (.775)			
	Group 6				.713 (.825)	.439 (.425)	.363 (.275)	.426 (.375)
Non-iterative: concurrent	Group 7	.840 (.925)	.791 (.925)	.737 (.875)	.657 (.775)			
	Group 8				.663 (.800)	.407 (.375)	.356 (.300)	.423 (.375)

We first tested whether TTO responses are robust to the procedural variations explored here. The full linear regression results are presented in Appendix 4 and summarised here. Recall that the chi tests here are testing for significant differences between models with and without these variables added i.e. between a full and reduced model. We find that the chi squared tests of differences for a model including the variants and their interactions were statistically significant (e.g. non-iterative sequential chi2(21) = 39.07 Prob > chi2 = 0.0096).
The results also showed that the value of 13122 (the common health state which everyone valued, and which is the intercept term in the regression) was not significantly different between the odd and even groups, indicating that TTO valuations are robust to 'context' effects (i.e. the coefficient on the Even dummy was not statistically different in the model presented in Appendix 4). As a sensitivity analysis, we reran the regression excluding responses that violate dominance and found similar results.

We also tested whether direct choices may be predicted from TTO responses assuming MUI and CPTO by exploring how well the aggregate choices coincide with TTO responses. Recall that if TTO and choice coincide we expect the 'splits' in the choices to be 50:50 on average and, hence the probability of choosing Life A or Life B to be 0.5. As above, no direct choice would be generated when the respondent valued a state as bad as dead or gave inconsistent responses within a non-iterative procedure such that no utility value may be estimated. This resulted in a fairly large number of respondents omitted from the choices particularly for those involving the more severe states and using a non-iterative procedure but we still had groups of at least 100 subjects. Tables 3 and 4 show the pairs of health states involved in the direct choices and number of respondents who answered each question for the odd and even versions respectively. Recall that one health state in the pair always appeared in Life A or Life B each time- in Tables 3 and 4 the health state that appeared in Life A is always the first in the pair. In each case, the direct choice was set up such that it was predicted (from their TTO responses) that the respondent would be indifferent between Life A and Life B- and, hence, we would expect a 50:50 on aggregate. We report first the overall splits of a preference for Life A (always involving the first health state in the pair) and Life B (always involving the second health state in the pair) in Tables 3 and 4 for the odd and even groups respectively.

Table 3: The main 'splits' of preferences for Lives A and B the direct choices in the odd groups (prediction was 50:50)

		Pairs of health states involved in the direct choices					
TTO variant	**Group**	**11121 vs 21211**	**11121 vs 12212**	**11121 vs 13122**	**21211 vs 12212**	**21211 vs 13122**	**12212 vs 13122**
Iterative: sequential (traditional TTO)	Group 1	69:31 (N=172) .000*	53:47 (N=168) .396	75:25 (N=165) .000*	52:48 (N=166) .103	72:28 (N=165) .000*	79:21 (N=163) .000*
Iterative: concurrent	Group 3	63:37 (N=179) .001*	64:36 (N=177) .000*	72:28 (N=170) .000*	62:38 (N=175) .290	73:27 (N=169) .000*	80:20 (N=168) .000*
Non-iterative: sequential	Group 5	63:37 (N=140) .003*	55:45 (N=146) .282	69:31 (N=134) .000*	52:48 (N=143) .0198	68:32 (N=151) .000*	79:21 (N=133) .000*
Non-iterative: concurrent	Group 7	69:31 (N=143) .000*	62:38 (N=138) .005*	77:23 (N=138) .000*	54:46 (N=138) .162	72:28 (N=148) .000*	76:24 (N=145) .000*

**One sample binomial test shows sig difference from p=0.5 (50:50 split) at 0.05 level of significance*

Table 4: The main 'splits' of preferences for Lives A and B the direct choices in the even groups (prediction was 50:50)

		Pairs of health states involved in the direct choices					
TTO variant	**Group**	**13122 vs 13224**	**13122 vs 23242**	**13122 vs 23314**	**13224 vs 23242**	**13224 vs 23314**	**23242 vs 23314**
Iterative: sequential (traditional TTO)	Group 2	64:36 (N=152) .001*	68:32 (N=158) .000*	67:31 (N=152) .000*	61:39 (N=145) .319	52:48 (N=151) .625	43:57 (N=145) .135
Iterative: concurrent	Group 4	65:35 (N=141) .001*	64:36 (N=132) .001*	64:36 (N=135) .001*	57:43 (N=123) .471	73:27 (N=130) .000*	55:45 (N=118) .311
Non-iterative: sequential	Group 6	63:37 (N=120) .005*	61:39 (N=126) .016*	64:36 (N=118) .002*	50:50 (N=121) .585	45:55 (N=131) .294	53:47 (N=112) .637
Non-iterative: concurrent	Group 8	57:43 (N=106) .025*	59:41 (N=104) .062	65:35 (N=107) .000*	56:44 (N=107) .082	42:58 (N=112) .299	40:60 (N=101) .111

**One sample binomial test shows sig difference from p=0.5 (50:50 split) at 0.05 level of significance*

The number of respondents answering each direct choice question ranged from 101 to 179 with the number greater in the odd versions using an iterative procedure (groups 1 and 3). The combination of the inconsistent responses in the non- iterative procedures and valuing at least one of the health states as worse than dead resulted in fewer respondents in the even non iterative versions of TTO (groups 6 and 8) being presented with the direct choice question. It is immediately obvious that many of the splits are a long way indeed from 50:50, with the most extreme split for 12212/13122 pairing being 80:20 in Group 3. The standard DCE approach would then assume that difference in utility between two lives was very large indeed-yet they have been set here to be equivalent since duration was chosen in order to produce two health profile with the same utility.

Tables 3 and 4 also show the results of the one- sample binomial tests of whether the probability of choosing Life A was significantly different than 0.5. It can be seen that, in the case of the traditional TTO, the null hypothesis (that p = 0.5) is rejected in the case of 4 and 3 of the 6 choices in Groups 1 and 2 respectively. In the case of the odd groups, none of the other variants resulted in the null hypothesis being accepted more often than in the traditional TTO. In the case of the even groups, the null hypothesis was accepted in 5 of the 6 choices, indicating that there may have been some slight tendency there to bring the TTO and choice closer, although caution has to be applied due to the smaller numbers of respondents in that group. A general trend shown in Table 4 is that the splits in the last 3 columns involving the more severe states are closer to 50:50. We return to this issue in the discussion.

It is obvious that the general tendency is to favour Life A that involves the first of the two health states reported in each pair and which more respondents valued higher than the other in TTO- than vice versa. We return to this issue below. In order to explore further the overall pattern across TTO variants, we combined the data across the 6 choices and used a GEE model to explore the extent to which choices differed significantly by the TTO method used to elicit the responses. In the GEE model the dependent variable was the probability of choosing Life B and the constant term estimated the impact on choice for the 'traditional TTO' in the odd group. As the constant terms shows a strong pattern to choose Life A that resulted in the move away from 50:50 in the splits, a positive coefficient on the other variables here is indicating that choices are closer to a 50:50 split than in the case of traditional TTO. Dummy variables and interaction effects were included to investigate whether the modelled effects were less pronounced in certain variants. For example, were the choices of respondents who had completed the 'traditional TTO' in the Even group (on the more severe states) significantly different than the choices of those who had completed the 'traditional TTO' in the Odd group (on the less severe states)? This comparison is captured by the dummy term 'even group =1'. Were the choices of those completing the 'non-iterative sequential' TTO on the odd states different than the choices of those completing the 'traditional TTO' on those same states? This comparison is captured by the 'non-iterative: sequential' dummy. Interaction terms were also included to investigate whether the impact of TTO variant differed across the Odd and Even groups.

Table 5: *The GEE model with 'Prob of choosing Life B' as dependent variable.*

TTO Group	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]
Constant	-0.6939	0.0855	-8.12	0.000	-0.8614 to -0.5264
Dummies					
Even Group (=1)	0.3075	0.1211	2.54	0.011	0.0701 to 0.5445
Iterative: concurrent	-0.1001	0.1208	-0.83	0.407	-0.3366 to 0.1365
Non-iterative: sequential	-0.0984	0.1245	0.79	0.429	-0.1457 to 0.3425
Non-iterative: concurrent	-0.0670	0.1266	-0.53	0.597	-0.3152 to 0.1812
Interactions					
Iterative:concurrent & Even	-0.0657	0.1749	-0.38	0.707	-0.4085 to 0.2770
Non-iterative:sequential & Even	0.0616	0.1777	0.35	0.729	-0.2867 to 0.4099
Non-iterative: concurrent & Even	0.1332	0.1832	0.73	0.467	-0.2259 to 0.4923
Wald chi2(5) = 140.96		Scale parameter: 1		Prob > chi2 = 0.0000	

**Based on 9677 observations.*

Table 5 shows the significant and positive coefficient on the 'Even' group dummy indicates a higher probability of choosing Life B in those groups- which results in choices that are closer to 50:50 than in the odd groups (given that the move away from 50:50 is in the direction of a preference for Life A).

We return to the issue of the general preference for Life A in the choices as it is worth looking at this in more detail. Recall that the health states that appeared under lives A and B were set in advance by the researchers. Thus, depending on the respondent's TTO valuations, either Life A or Life B could involve fewer years in the 'better' health state or vice versa. In order to explore the pattern of choices in terms of whether the respondent was selecting the Life involving fewer years in a better state-or vice versa- Tables 6 and 7 shows the relationship between direct choice and the respondent's own TTO values according to whether, U1>U2, U1=U2 or U1<U2 in each pair. When U1> U2, Life A would then involve the shorter time in the better (for that respondent) health state. When U1=U2, Lives A and B would involve the same number of life years and when U1 < U2, Life A would involve the longer time in the worse (for that respondent) health state. Tables 6 and 7 shows in brackets the number of respondents with each TTO pattern and the percentage of those respondents that went onto choose Life A in direct choices. So, for example, for the 11121 vs 21211 comparison in Group1, 60 respondents valued 11121 more highly than 21211 in TTO and 47% of those went onto choose Life A involving fewer years in state 11121. Seventy six respondents valued 11121 equal to 21211 and 76% of those went onto choose Life A in direct choice- in this case involving the same number of years life in 11121 and 21211.

We begin by looking at the cases of dominance where clear predictions may be made (11121 dominates 13122 in the odd groups and 13122 dominates 13224 and 23242 in the even groups). Where one state dominates another, but receives the same valuation in TTO, we would expect respondents to overwhelmingly choose the dominant state in a straight choice involving the same number of life years and our data confirms this finding.

What is more interesting, however, is that even when no dominance exists and U1=U2 in the TTO, there is often a strong preference for Life A involving the first health state in the pair. This cannot be explained by a preference for a shorter time in a better health state- or vice versa- as the life years in the direct choices are then equivalent. What we have uncovered appears to be a strong preference for the life involving the first state in the pair- which the *majority* of respondents who had made a distinction in TTO had valued more highly than the other. This suggests that at least a number of respondents did agree with the aggregate ranking of the health states, but that was not reflected in their TTO responses. We return to this issue in the discussion.

Table 6: The percentage of respondents choosing 'Life A' involving the first state in each pair broken down by respondents own TTO responses (numbers in each category)-Odd groups

		Pairs of health states involved in the direct choices					
TTO method	**TTO values***	**11121 vs 21211**	**11121 vs 12212**	**11121 vs 13122**	**21211 vs 12212**	**21211 vs 13122**	**12212 vs 13122**
Group 1 **Iterative: sequential (traditional TTO)**	U1> U2	47% *(60)*	40% *(93)*	66%*(104)*	39% *(88)*	62% *(96)*	64% *(73)*
	U1 = U2	76% *(76)*	68% *(53)*	91% *(45)*	60% *(53)*	87% *(46)*	95% *(37)*
	U1 < U2	89% *(36)*	76% *(21)*	81% *(16)*	79% *(24)*	72% (23)	89% *(52)*
Group3 **Iterative: concurrent**	U1> U2	39% *(83)*	49% *(112)*	62%*(120)*	41% *(93)*	64% *(108)*	65% *(80)*
	U1 = U2	80% *(64)*	87% *(53)*	97% *(34)*	83% *(46)*	88% *(40)*	88% *(42)*
	U1 < U2	91% *(32)*	100% *(12)*	100%*(16)*	89% *(36)*	95% *(21)*	98% *(46)*
Group 5 **Non-iterative: sequential**	U1> U2	42% *(62)*	52% *(90)*	63%*(104)*	34% *(67)*	60% *(98)*	66% *(67)*
	U1 = U2	80% *(64)*	48% *(44)*	96% *(27)*	*60% (52)*	83% *(42)*	94% *(47)*
	U1 < U2	79% *(14)*	100% *(12)*	67% *(3)*	88% *(24)*	73% *(11)*	90% *(19)*
Group 7 **Non-iterative: concurrent**	U1> U2	59% *(93)*	54% *(68)*	69% *(94)*	41% *(68)*	66% *(90)*	68% *(72)*
	U1 = U2	71% *(80)*	66% *(53)*	93% *(41)*	54% *(61)*	84% *(50)*	88% *(49)*
	U1 < U2	88% *(17)*	82% *(17)*	100% *(3)*	90%*(19)*	80% *(5)*	79% *(14)*

**where U1 refers to the respondent's own TTO valuation for the first health state in each pair and U2 to the respondent's own TTO valuation for the second health state in the pair.*

Table 7: The percentage of respondents choosing 'Life A' involving the first state in the pair broken down by respondents own TTO responses (numbers in each category)-Even groups

		Pairs of health states involved in the direct choices					
TTO method	TTO values *	**13122 vs 13224**	**13122 vs 23242**	**13122 vs 23314**	**13224 vs 23242**	**13224 vs 23314**	**23314 vs 23242**
Group 2 Iterative: sequential (traditional TTO)	U1> U2	59% *(112)*	66% *(128)*	34%*(106)*	56% *(70)*	41% *(59)*	45% *(67)*
	U1 = U2	70% *(20)*	78% *(18)*	72% *(25)*	52% *(25)*	49% *(49)*	62% *(24)*
	U1 < U2	85% *(20)*	75% *(12)*	81% *(21)*	72% *(50)*	72% *(43)*	70% *(53)*
Group 4 Iterative: concurrent	U1> U2	57% *(105)*	59% *(106)*	38% *(98)*	46% *(66)*	51% *(49)*	40% *(58)*
	U1 = U2	69% *(13)*	91% *(11)*	81% *(16)*	73% *(22)*	73% *(37)*	42% *(26)*
	U1 < U2	96% *(23)*	87% *(15)*	91% *(21)*	69% *(35)*	98% *(44)*	66% *(34)*
Group 6 Non-iterative: sequential	U1> U2	57% *(93)*	58% *(107)*	59% *(95)*	46% *(59)*	26% *(51)*	37% *(48)*
	U1 = U2	85% *(20)*	67% *(12)*	80% *(15)*	52% *(25)*	48% *(50)*	61% *(26)*
	U1 < U2	86% *(7)*	100% *(7)*	100% *(8)*	54% *(37)*	73% *(30)*	63% *(38)*
Group 8 Non-iterative: concurrent	U1> U2	55% *(85)*	56% *(86)*	67% *(81)*	40% *(48)*	22% *(23)*	48% *(52)*
	U1 = U2	86% *(14)*	80% *(15)*	76% *(21)*	63% *(27)*	59% *(54)*	59% *(22)*
	U1 < U2	86% *(7)*	67% *(3)*	100% *(5)*	75% *(32)*	71% *(35)*	78% *(27)*

**where U1 refers to the respondent's own TTO valuation for the first health state in each pair and U2 to the respondent's own TTO valuation for the second health state in the pair.*

DISCUSSION

We systematically varied aspects of TTO in order to bring TTO more in line with how choices would be presented in a DCE that set out to derive utility values using time as the numeraire-sometimes referred to as DCE_{TTO}. We found that TTO responses were not robust to the procedural variations tested here, which is similar to previous studies that have found that different procedures yield different results[13,14]. For example, it has previously been shown that the elicitation procedures used[15], whether 'props' are used or not[16] and the mode of administration of the survey[17] can all affect the values derived. We then tested whether it was possible to use the TTO valuations to predict direct choices between health states. The direct choices were set up such that any individual respondent ought to be indifferent between the lives on offer in the direct choice and, hence, there would be a 50:50 split in aggregate. We found that a number of the splits were a long way indeed from 50:50, but that the divergence between TTO and direct choice did not disappear when alternative TTO variants were deployed. The divergence from 50:50 was not, however, random but systematically favoured Life A which always involved the state that the *majority* of respondents (who had made a distinction) in the TTO had valued more highly. Those respondents who valued two states equally in the TTO (and, hence, were presented with direct choices involving the same number of life years) overwhelmingly went for the life involving the state that the *majority* had rated as better. This effect was more marked in the odd groups involving the less severe states.

Some of the results do appear relatively easy to explain. Respondents who valued two states equally in the TTO and, hence, were faced with the same number of life years in the direct choice, often had a strong aggregate preference for one state over the other. Based on how DCE responses are analysed, the utility values of those states would be assumed to be very far apart- and yet they were valued equally in the TTO. At least some of this anomaly may be explained by the fact that respondents who did not trade life years in the TTO were presented with the same durations in the direct choice. Even a slight preference for one health state over the other would then lead them to choose the life involving that state without having to sacrifice any life expectancy. Whilst there is no a priori reason to suppose, for example, that 11121 is better than 21211 for any particular respondent, it appears that many who did value the states equally in TTO *did* consider 11121 to be better than 21211 and that preference came out in the direct choice. But only around half of respondents who valued 11121 equal to 21211 in the TTO were non traders on both states, so non-trading alone cannot explain the pattern fully.

In setting up the choices we did not assume the linear QALY model to hold as we believed that to be too restrictive. Whilst our approach did allow for more flexibility than under the linear QALY model which is often deployed elsewhere, it could be argued that even the assumptions of MUI and CPTO are strong ones. Whilst space did not permit a detailed discussion of the method here, we did run a parallel study in which the direct choices were set up in such a way that relied only on transitivity. The data showed that failures of the assumptions of MUI and CPTO were *not* the main drivers of the results reported here. Details of the parallel study are available from the authors on request.

We do, however, believe there may be an explanation of our findings in terms of the differential error structure of TTO and choice. Suppose that in TTO subjects maximize utility functions U(11121)=V(11121)+error and U(21211)=V(21211)+error and V(11121) > V(21211), where U(.) is the utility used by the respondent in the TTO questions and V(.) is the 'true' utility value for this subject. Overall, the majority of cases will state U(11121)>U(21211) when V(11121)>V(21211). But there will be some respondents, for whom V(11121)>V(21211) and yet who stated U(11121)<U(21211) due to the overlapping nature of utility distributions. However, in a direct choice it can be easier for these people to observe that (11121) is milder than (21211) as only ordinal preferences are required.

Of course, the choice results may be driven not only by differences in intrinsic utility but also by how easy it is for respondents to see that one state is better than another, termed 'comparability'[18]. Likewise, Tversky[19] noted that "choice probabilities, therefore, reflect not only the utilities of the alternatives in question, but also the difficulty of comparing them" (p. 284) and for this reason "the probability of selecting an alternative depends not only on its overall value, but also on its relations to the other available alternatives" (p. 295). This led him to question the assumption that choices can be represented by independent random variables, that is, by an independent random utility model. Our results raise the possibility that something like this that may be happening in choices between health profiles. One possible explanation of the finding that the splits were closer to 50:50 for the more severe states is that they are more difficult to compare in that more levels and dimensions are changing at one time. Further discussion of these issues is beyond the scope of the current paper, but it raises important questions about the fundamental assumptions underpinning most DCE models if it is true that choices may be driven not only by differences in utilities but also by how *easy* it is to compare alternatives. This would generate changes in the error structure of the model used to link choices and utility.

A limitation of the study is that health states were not randomised to Life A and Life B in the direct choice, and the strong preference for Life A could, of course, indicate a tendency to favour the left hand option. This explanation cannot, however, account for all the findings as it cannot explain the different patterns observed between Odd and Even groups.

It then seems likely that the disparity between TTO and direct choice that we find in our paper is being driven by a combination of factors, certain to do with ‘problems’ with TTO that are already well known about (such as insensitivity and non-trading for mild health states) whilst others are to do with the appropriate interpretation of choice data which has been rather less explored to date. It would appear though that the combination of these factors are more important drivers of the disparity between TTO and choice than the procedural issues we set out to look at here. We recommend that future research address the issue of choices being driven by factors other than differences in utilities and for this to be explored in a systematic way.

Acknowledgements

We would like to thank Ruben Mujica-Mota for his help and advice on the use of Generalised Estimating Equations in this dataset. The study was funded by the Spanish Ministry of Education, Research Grant ECO2010.22041.C02.01 and Junta de Andalucía, Project P09.SEJ.4992. The study received ethics approval at the Research Ethics committee at Glasgow Caledonian University.

Anne Spencer was supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South West Peninsula. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

REFERENCES

1. Bansback, N., J. Brazier, A. Tsuchiya and A. Anis (2012). "Using a discrete choice experiment to estimate health state utility values." J Health Econ **31**(1): 306-318.
2. Bansback, N., A. R. Hole, B. Mulhern and A. Tsuchiya (2014). "Testing a discrete choice experiment including duration to value health states for large descriptive systems: addressing design and sampling issues." Soc Sci Med **114**: 38-48.
3. Brazier, J., D. Rowen, Y. L. Yang and A. Tsuchiya (2012). "Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale." European Journal of Health Economics **13**(5): 575-587.
4. Norman, R., R. Viney, J. Brazier, L. Burgess, P. Cronin, M. King, J. Ratcliffe and D. Street (2013). "Valuing SF-6D Health States Using a Discrete Choice Experiment." Med Decis Making.
5. Ratcliffe, J., L. Couzner, T. Flynn, M. Sawyer, K. Stevens, J. Brazier and L. Burgess (2011). "Valuing Child Health Utility 9D health states with a young adolescent sample: a feasibility study to compare best-worst scaling discrete-choice experiment, standard gamble and time trade-off methods." Applied Health Economics and Health Policy **9**(1): 15(13)
6. Stolk, E. A., M. Oppe, L. Scalone and P. F. M. Krabbe (2010). "Discrete Choice Modeling for the Quantification of Health States: The Case of the EQ-5D." Value in Health **13**(8): 1005-1013.
7. Tversky, A., S. Sattath and P. Slovic (1988). "Contingent weighing in judgment and choice." Psychological Review **95**: 371-384.
8. Stalmeier, P. Wakker, P and Bezembinder T (1997) “Preference reversals: Violations of unidimensional procedure invariance”, Journal of Experimental Psychology **23** (4) 1196-1205.
9. Sumner, W. and Nease R.F (2001). “Choice-Matching Preference Reversals in Health Outcome Assessments”, Medical Decision Making 21: 208-218
10. Attema, A and Brouwer B.F (2013), “In search of a preferred preference elicitation method: A test of the internal consistency of choice and matching tasks” Journal of Economic Psychology**, 39** 126-140
11. Delquie, P. (1993). "Inconsistent Trade-Offs between Attributes - New Evidence in Preference Assessment Biases." Management Science **39**(11): 1382-1395.
12. Fischer, G. W., Z. Carmon, D. Ariely and G. Zauberman (1999). "Goal-based construction of preferences: Task goals and the prominence effect." Management Science **45**
13. Arnesen, T. and M. Trommald (2005). "Are QALYs based on time trade-off comparable?--A systematic review of TTO methodologies." Health Econ **14**(1): 39-53.
14. Attema, A. E., Y. Edelaar-Peeters, M. M. Versteegh and E. A. Stolk (2013). "Time trade-off: one methodology, different methods." European Journal of Health Economics **14**: S53-S64.
15. Lenert, L.and G. Alan (1998). "The effect of search procedures on utility elicitations." Medical Decision Making **18**: 76-83.
16. Dolan, P., C. Gudex, P. Kind and A. Williams (1996). "Valuing health states: a comparison of methods." J Health Econ **15**(2): 209-231.
17. Norman, R., M. T. King, D. Clarke, R. Viney, P. Cronin and D. Street (2010). "Does mode of administration matter? Comparison of online and face-to-face administration of a time trade-off task." Qual Life Res **19**(4): 499-508.
18. Krantz, D. H. (1967). "Extensive Measurement in Semiorders." Philosophy of Science **34**: 348-362.

19. Tversky, A. (1972). "Elimination by Aspects: A Theory of Choice" Psychological Review 7**9**: 281-299.

Appendix 1: *The assumptions used in setting up the direct choices.*
We illustrate the conceptual framework with some simple notation. Suppose that in the TTO method each alternative is characterized by a pair (q,t) where q is quality of life and t is time. Consider now the quality of life of health state i, denoted by q_i, and for simplicity let us assume that this health state is strictly worse than normal health. In order to establish indifference in the TTO, subjects have to undertake a sequence of binary choices (q_{NH}, t^*_{NH}) vs $\left(q_i, t_i\ \right)$ to determine in which interval the value of health state i lies, where normal health is q_{NH}. In the conventional TTO, t_i is kept constant and $t^*_{NH}$is adjusted until indifference is reached (where the asterisk is used to denote the variable that is adjusted to reach indifference). Assume that at indifference the subject sets $(q_{NH}, t^*_{NH}) \sim \left(q_i, t_i\ \right)$ where $t_i \succ t^*_{NH}$ since $q_i \prec q_{NH}$.

Consider now two such TTO questions, that are used to value health states 1 and 2 respectively (i.e. i=1, 2) and t_i is kept constant (i.e. $t_i = t$). We have then that $(q_{NH}, t^1_{NH}) \sim (q_1, t)$ and $(q_{NH}, t^2_{NH}) \sim (q_2, t)$ where and t^1_{NH} and t^2_{NH} are the durations that make the subject indifferent between for health states 1 and 2 respectively. The study then uses these two TTO judgments to present a choice comparing health state 1 and 2 directly. So for example, respondents are asked to compare t_1 in health state 1 and t_2 in health state 2. In the tests set out in this paper $(t_1\ , t_2\ )$ are set so that the respondent should be indifferent between the two alternatives, i.e. $\left(q_1, t_1\ \right) \sim \left(q_2, t_2\ \right)$ as explained in more detail below.

If mutual utility dependence holds, and the utility function is multiplicative then $(q_{NH}, t^1_{NH}) \sim (q_1, t)$ and $(q_{NH}, t^2_{NH}) \sim (q_2, t)$ should imply that U(q_{NH}) x U(t^1_{NH})=U(q_1) x U(t) and U(q_{NH}) x U(t^2_{NH})=U(q_2) x U(t) and then:

$$\frac{\mathrm{U}(q_{NH}) \text{ x } \mathrm{U}(t^1_{NH})}{\mathrm{U}(q_{NH}) \text{ x } \mathrm{U}(t^2_{NH})} = \frac{\mathrm{U}(q_1) \text{ x } \mathrm{U}(t)}{\mathrm{U}(q_2) \text{ x } \mathrm{U}(t)} \qquad (1)$$

And with the cancellation of common terms this expression simplifies to:

$$\frac{\mathrm{U}(t^1_{NH})}{\mathrm{U}(t^2_{NH})} = \frac{\mathrm{U}(q_1)}{\mathrm{U}(q_2)} \qquad (2)$$

Let us now choose any durations t_1 and t_2 such that:

$$\frac{t_1}{t^1_{NH}} = \frac{t_2}{t^2_{NH}} = k \qquad (3)$$

Rearranging question (3) gives $t^1_{NH} = \frac{t_1}{k}$ and $t^2_{NH} = \frac{t_2}{k}$ which we substitute into equation (2).

Constant proportional trade off (CPTO) assumes that the respondent will always trade off the same *proportion* of life years in the TTO. Under CPTO the utility function for life years, U(t), is homogeneous, which implies that U(K t) = K x U(t) where K is any constant term and t is time. So under CPTO $\mathrm{U}(t^1_{NH}) = \frac{\mathrm{U}(t^1_{NH})}{k}$. As this also holds for health state 2, then, we have with the cancellation of common terms that,

$$\frac{\mathrm{U}(t_1)}{\mathrm{U}(t_2)} = \frac{\mathrm{U}(q_1)}{\mathrm{U}(q_2)} \rightarrow U(t_1, q_2) = U(t_2, q_1)$$

So, from two TTO questions, we can select pairs (t_1 , t_2) such that we can generate choices between two health profiles of the same utility. At the aggregate level, we would expect half of the respondents choosing one option and half choosing the other option in a forced binary choice task. This is the main test we will conduct in this paper.

Appendix 2: Sections 1-3 of the survey

In section 1, respondents first saw a general introduction;

Thank you for agreeing to take part in this survey. The purpose of the survey is to try and find out what matters *to members of the public- like yourself- when it comes to thinking about health improvements. But, as you will be aware, there are thousands of different types of treatments that could be funded on the NHS and we cannot ask about them all. What we are going to do here instead is to ask about what matters in general to people in terms of health improvements. For example, some treatments improve quality of life, others prolong life expectancy, whilst others improve both the quality and length of life. This survey is going to ask you a number of questions designed to find out about the relative importance you place on different types of health improvements. There are no right or wrong answers- we just want to know what you personally think.*

Respondents were then presented with 3 statements in turn and asked to indicate the strength of their agreement on a 5 point Likert scale running from strongly agree (1) to strongly disagree (5):

1. I would always prefer to live as long as possible regardless of what my quality of life was.
2. I would always prefer to have a good quality of life than to live for a long time in a poor health state.
3. I would rather be dead than live in a really bad health state in which my quality of life was very low.

The purpose of this task was to get respondents thinking in general terms about quality and length of life before they were faced with the TTO questions. Respondents were then introduced to the EQ 5D (5L) descriptive system and asked, in turn, which of the 5 dimensions were the most and least important to them personally (ties were allowed). They were then asked to identify their own health state on the EQ 5D (5L) descriptive system.

In section 2 respondents were first asked to think about how 'good' or 'bad' it would be to be in a particular health state and were introduced, in turn, to EQ 5D (5L) states 11111 (hereafter labelled 'normal health'), 12111, 23322 and 43545. The last 3 health states would not feature in the TTO exercises to follow, but were included here in order to provide a common 'frame' to all respondents. In particular, state 43545 was more severe than any that would feature in the exercises to follow, but we were keen to introduce all respondents to wide range of EQ 5D states.

In section 3 respondents were introduced to the *general* idea behind TTO exercises. Respondents were first asked to consider being in state 23322 for 20 years after which time they would die- denoted by 'Life A' They were asked to think about being in the 'good health' state for 20 years – denoted by 'Life B'. They were told that 'in this case Life B is clearly better than Life A' and were then asked 'but what if life B was fewer than 20 years- what if life B was 13 years?' They were then asked just to think about which of the two lives they would prefer - but without having to record a response. They were then told that 'in the screens that follow', they would be choosing between a different Lives A and B each time.

Appendix 3: Achieved sample

	Sample	UK		Sample	UK
Male 18-25	3.7%	6.4%	**Female 18-25**	5.0%	6.4%
Male 26-35	10.3%	8.8%	**Female 26-35**	12.3%	9.6%
Male 36-45	9.1%	9.6%	**Female 36-45**	10.0%	9.6%
Male 46-54	11.2%	7.2%	**Female 46-54**	13.0%	8.0%
Male 55-64	7.4%	6.4%	**Female 55-64**	8.8%	7.2%
Male 65+	5.1%	8.8%	**Female 65+**	4.1%	12.0%

Appendix 4: ***Linear regression of impact of TTO variant on TTO valuations.***

Variable	coefficient	Standard error		Coefficient	Standard error	
constant	0.6839	0.0164	*	0.6669	0.0232	*
State11121	0.1409	0.0179	*	0.1463	0.0187	*
State21211	0.1151	0.0179	*	0.1205	0.0187	*
State12212	-0.0192	0.1797		-0.0138	0.0188	
State13224	-0.2171	0.0179	*	-0.2225	0.0187	*
State23242	-0.2739	0.0179	*	-0.2793	0.0187	*
State23314	-0.2290	0.0179	*	-0.2344	0.0187	*
Methodic	-0.0536	0.0230	*	-0.0393	0.0326	
Methodnis	0.0048	0.0237		-0.0051	0.0333	
Methodnic	-0.0301	0.0241		-0.0238	0.0336	
ic_11121	0.0334	0.0252		0.0288	0.0264	
ic_21211	0.0182	0.0253		0.0136	0.0264	
ic_12212	0.0515	0.0253	*	0.0469	0.0265	
ic_13224	-0.0226	0.0252		-0.0179	0.0263	
ic_23242	-0.0437	0.0253		-0.0390	0.0264	
ic_23314	-0.0264	0.0253		-0.0218	0.0264	
nis_11121	0.0473	0.0265		0.0511	0.0276	
nis_21211	0.0233	0.0260		0.0270	0.0271	
nis_12212	0.0650	0.0264	*	0.0689	0.0276	*
nis_13224	-0.0517	0.0263	*	-0.0567	0.0276	*
nis_23242	-0.0555	0.0265	*	-0.0601	0.0277	*
nis_23314	-0.0626	0.0265	*	-0.0674	0.0278	*
nic_11121	0.0185	0.0267		0.0166	0.0278	
nic_21211	0.0231	0.0264		0.0212	0.0273	
nic_12212	0.0959	0.0265	*	0.0941	0.0276	*
nic_13224	-0.0360	0.0267		-0.0347	0.0281	
nic_23242	-0.0186	0.0269		-0.0172	0.0282	
nic_23314	-0.0145	0.0268		-0.0132	0.0282	
GROUP_EVEN				0.0330	0.0328	
ic_EVEN				-0.0286	0.0460	
nis_EVEN				0.0210	0.0473	
nic_EVEN				-0.0116	0.0481	
Overall r-squared	0.2500			0.2242		

*statistically significant at a 0.05 level

We show two regressions: the first looks at the impact of variants upon the TTO values and the second additionally investigates whether the value of 13122 (the common health state which everyone valued, and which is the intercept term in the regression) was significantly different between the odd and even groups.

Definition of variables that appear in the model reported in Appendix 4.

constant = value of health state 13122 under the traditional TTO variant. For the regression that excludes dummies for odd and even groups, the constant represents the value of health state 13122 averaged across odd and even groups. In the fifth and sixth column of the table the constant is the value of health state 13122 under the traditional TTO variant for the odd group.

State11121, State21211, State12212, State13224, State23242 and State23314 are dummies for the additional impact on utilities (compared to 13122) of the associated EQ 5D state.

methodic is the impact of the iterative concurrent variant upon the utility value of state 13122
methodnis is the impact of non-iterative sequential variant upon the utility value of state 13122
methodnic is the impact of non-iterative concurrent variant upon the utility value of state13122

The interaction terms look at whether the variants affected the valuations of the other health states, e.g.
ic_11121 looks at the impact of the iterative concurrent variant on state 11121.
nis_11121 looks at the impact of the non-iterative sequential variant on state 11121.
nic_11121 looks at the impact of the non-iterative concurrent variant on state 11121.
Likewise, for the remaining health states.

GROUP_EVEN is whether state 13122 was valued differently in the even groups than the odd groups in the traditional TTO.

Further interaction terms e.g. ic_EVEN look at whether health state 13122 was valued differently for the method iterative concurrent in the Even groups. Similar dummies are defined for the other variants.