HEALTH ECONOMICS GROUP

Faculty of Medicine and Health
Norwich Medical School

University of East Anglia

Exploring differences between TTO and direct choice in the valuation of health states

Angela Robinson[1], Anne Spencer[2], Jose-Luis Pinto-Prades[3], Judith Covey[4]

[1]Health Economics Group, Norwich Medical School, University of East Anglia, Norwich UK

[2]Exeter Medical School, University of Exeter, Exeter UK

[3]Yunus Centre for Social Business & Health and Glasgow School for Business & Society, Glasgow Caledonian Unirsity, Glasgow UK

[4]Department of Psychology, Durham university, Queen's Campus, Stockton on Tees UK

Exploring differences between TTO and direct choice 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,
Salmon Pool Lane,
Exeter,
EX2 4SG
a.e.spencer@exeter.ac.uk

c. Yunus Centre for Social Business & Health and Glasgow School *for* Business & Society,
Glasgow Caledonian University,
58 Port Dundas Road.
Glasgow
G4 0BA
joseluis.pinto@gcu.ac.uk

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

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 choice. We designed an interactive survey that was hosted on the internet and 2022 interviews were achieved in the UK. We found a marked divergence between TTO and choice, but this was not related to the 'transparency' of the TTO procedure. We conclude that a combination of insensitivity in the TTO (however conducted) and factors other than differences in utility affecting choices is driving the divergence.

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

INTRODUCTION

There has been recent interest in the use of discrete choice experiments (DCE) in utility values for health states valuing health states health state elicitation using the DCE method to derive health state utilities for use in QALY calculations (Ratcliffe, Couzner et al. 2011, Bansback, Brazier et al. 2012, Brazier, Rowen et al. 2012, Norman, Viney et al. 2013, Bansback, Hole et al. 2014). When compared head to head, DCE and TTO have been shown to arrive at different utility estimates (Stolk, Oppe et al. 2010, Brazier, Rowen et al. 2012) but to date little research has gone into exploring the factors that might be driving these differences. The aim of this study is address this gap in the literature.

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 (Ratcliffe, Couzner et al. 2011, Norman, Viney et al. 2013). For example, in discussing the TTO, Bansback and colleagues(Bansback, Brazier et al. 2012) 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 is subsequently modelled. Even leaving the functional form of the model aside, there may be other reasons why a DCE that sets out to value health state utilities using time as the numeriare –sometimes referred to as DCE_{TTO}[1] - may lead to systematic differences in valuations than traditional TTO. In this paper we draw on the psychological literature with the aim of developing a framework within which to explain why TTO and choice data may differ.

It has been observed for some time that preferences between two options can change depending on the elicitation procedure used, a phenomena known as preference reversals (Lichtens.S and

[1] We are distinguishing DCE_{TTO} from a DCE that sets out to value health states by using risk as a numeriare- which would more closely resemble the SG method.

Slovic 1971). In particular, it is well known that 'matching' and 'choice' tasks yield different results (Tversky, Sattath et al. 1988). 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 (Tversky, Sattath et al. 1988, Delquie 1993). 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 hone in on a point of indifference, a technique which has been referred to as 'choice-based matching'(Fischer, Carmon et al. 1999). Fischer et al 1999 developed the task-goal hypothesis and argued that more weight is given to the prominent attribute when the aim is to differentiate amongst options, as in choice, than in tasks where the aim is to equate options, as in matching. They then went on to show when the objective – or goal – of the matching task was made less transparent, using such as methods as 'hidden choice-based matching' (HCBM) (Fischer, Carmon et al. 1999) preference reversals between matching and choice were reduced.

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, given the findings of Fischer and others, we hypothesise this would reduce any differences that may exist 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.

STUDY OBJECTIVES

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 the quality of life and t is time. In order to establish indifference for health state B subjects have to undertake a sequence of binary choices (q_{FH}, t_{NH}) vs (q_B, t_B), to determine in which interval the values lie, where normal health is q_{NH}. Since $q_B \prec q_{NH}$ indifference requires $t_B \succ t_{NH}$.

The issue of interest here is whether we can predict choices between health profiles based on the respondent's TTO valuations? Assume that in TTO a respondent sets $(q_{NH}, t_{NH}^B) \sim (q_B, t_B^*)$ and $(q_{NH}, t_{NH}^C) \sim (q_c, t_C^*)$. Can we predict how the respondent will choose between $\left(q_B, t_B\right)$ and $\left(q_C, t_C\right)$? 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 in which case we simply estimate the total number of QALYs of each alternative in choice and predict that the respondent will choose the alternative offering the greater higher number of QALYs. Linearity is, however, a very restrictive assumption, so we rely here on the weaker conditions of utility independence and constant proportional trade-off which allow subjects to discount future health.

Assume that $(q_{NH}, t_{NH}^c) \sim (q_C, t^*)$ and $(q_{NH}, t_{NH}^B) \sim (q_B, t^*)$. Where NH stands for Normal Health, A and B are health states worse than NH, t^* is the duration in bad health and t_{NH}^c and t_{NH}^B are the durations that make the subject indifferent between t^* years in bad health and t_{NH}^c and t_{NH}^B years in NH for health states C and B respectively. From these two TTO judgments we want to generate a choice between $\left(q_B, t^B\right)$ and $\left(q_C, t^C\right)$ such that $\left(q_B, t^B\right) \sim \left(q_C, t^C\right)$.

If the utility function is multiplicative then U(q_{NH}) x U(t_{NH}^c)=U(q_C) x U(t^*) and U(q_{NH}) x U(t_{NH}^B)=U(q_B) x U(t^*) and then:

$$\frac{\mathrm{U}(q_{NH}) \text{ x } \mathrm{U}(t_{NH}^c)}{\mathrm{U}(q_{NH}) \text{ x } \mathrm{U}(t_{NH}^B)} = \frac{\mathrm{U}(q_C) \text{ x } \mathrm{U}(t^*)}{\mathrm{U}(q_B) \text{ x } \mathrm{U}(t^*)}$$

Let us now choose any durations t^B and t^C such that

$$\frac{t^B}{t_{NH}^B} = \frac{t^C}{t_{NH}^C} = k$$

Let us now assume that CPTO holds. Then U(t^B)=kxU(t^B_{NH})[2] and the same applies to C. Then, we have that,

$$\frac{\mathrm{U}(t^c)}{\mathrm{U}(t^B)} = \frac{\mathrm{U}(q_C)}{\mathrm{U}(q_B)} \rightarrow U(t^c, q_B) = U(t^B, q_C)$$

So, from two TTO questions, we can select pairs (t^B, t^C) such that we can generate choices between two health profiles with the same utility. At the aggregate level, we would expect half of the subjects 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.

Of course, even the assumptions of Utility Independence and CPTO are strong ones and there is evidence that they may not hold. Hence, we also include a test that does not rely on the basic TTO model being correct. Suppose that, unlike in the conventional TTO, we keep t_{NH} constant and adjust t_B until indifference is reached. If t_B^* is the level of t_B such that $(q_{NH}, t_{NH}) \sim (q_B, t_B^*)$, then we use t_{NH} in order to estimate the indifference point between (q_{NH}, t_{NH}) and (q_C, t_C) and $(q_{NH}, t_{NH}) \sim (q_B, t_C^*)$. It should then hold, assuming transitivity alone, that $(q_B, t_B^*) \sim (q_c, t_C^*)$. We test this in this study and refer to the method as 'reverse TTO' which is explained in detail below. Hence, we set out to test for convergence between TTO and direct choice both when the traditional TTO assumptions are considered to hold- and when they are relaxed.

In the case where Utility independence and CPTO *are* assumed to hold, we 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:

1. Whether an iterative or non-iterative procedure is used to arrive at a point of indifference.

2. 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 that to; a) systematically influence the TTO valuations

[2] CPTO implies that the utility function for life years is homogeneous, that is, U(k Y) = k x U(Y)

themselves and b) bring about convergence between TTO and direct choice. The objectives of the current study were 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 assuming Utility Independence and CPTO.
3. Examine to what extent direct choices may be predicted from TTO responses when Utility Independence and CPTO are relaxed.

METHODS

Survey design

In order to explore all factors of interest, but without over burdening respondents, 9 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 1. In section 4, respondents were randomised to one of 9 versions of the survey according to which variant of TTO – and set of health states- they would see. Local ethics approval was granted by Glasgow Caledonian University.

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 versions 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 in order to include states of 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 in the set. 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 3 ‘iterative’ variants in detail.

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

This variant replicates a ‘traditional’ TTO exercise. 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 the ‘switched’ to preferring Life A to Life B in successive two year intervals –when they were then asked about the year in between. For example, if the ‘switched’ from preferring Life A to Life B – or vice versa – 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[3]. 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 have a value of ‘1’ for that health state.

[3]This is the same level of accuracy as the ‘something and 6 month’ question in a 10 year TTO.

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[4].

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	(shaded)	NO problems in walking about	
NO problems washing or dressing oneself		NO problems washing or dressing oneself	
Slight problems doing usual activities	(shaded)	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

The procedure described thus far is exactly the same as that used in the other iterative TTO versions described below. The important point here, however, is that in the traditional TTO, the iterative procedure is followed through to the end for one health state, before moving on to the next health state. 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

[4] This was for pragmatic reasons in not wanting to complicate the survey further, but will not impact on any of the tests carried out.

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 akin to the ‘hidden choice-based matching’ approach 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 take place concurrently. Hence, even within an iterative procedure, it is less ‘transparent’ to respondents what their task is in relation to any particular 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 ‘hone in’ on a point of indifference. Rather, respondents are presented with choices that are not generally based on their previous responses. It could be argued that the use of a non-iterative TTO procedure is again making it less transparent to respondents that the 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-sequentially’ versions 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[5]. 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, in this version the valuation procedure is followed all the way through for one 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, hence observations may be missing.

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. 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. Thus, for example, the first 16 questions respondents in the odd group were presented with were a random draw from the table shown in Appendix 2. 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 described in the ‘sequential’ version described above–and ended after the 4 health states had been valued.

[5] This is because a wholly non-iterative system that assessed utility values to the same degree of accuracy as in the traditional TTO variant above, would entail presenting respondents with 20 choices for each health state. We considered that to be too many, so the approach used here is more accurately described as ‘semi iterative’.

Table 1 shows the first 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** *	**Group2** *	**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 in another. 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.

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. These life expectancies were chosen as we did not want respondents anchoring on a 20 year life expectancy as in the TTO exercises, but we were keen to avoid reducing life expectancy too much as that may introduce other 'framing' effects. 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. Those respondents whose TTO valuations of the two states were in the opposite direction 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.

The 'special case': the 'reverse TTO' and associated direct choice questions

As above, we were keen to test whether TTO and choice coincides without applying strong assumption of constant proportional trade off. Respondents randomised to the 'reverse TTO' are first presented with a choice between 10 years in Life A and 5 years in Life B. The scenario as presented to respondents is depicted in Appendix 3 using state 21211 as an example. Depending upon their choice, the time in good health was iterated up or down and the iterative process continued until the 'switched' to preferring Life A to Life B in successive two year intervals. As the 'reverse TTO' cannot be applied to states worse than dead – since additional time in those states reduces utility-we only tested the 'reverse TTO' in the odd groups where the health states are generally mild.

Suppose that V1 and V2 each represent the number of years in states 1 and 2 respectively that made the respondent indifferent between that life and 5 years in normal health. In the 'reverse

TTO' respondents would then be presented with a choice between Life A: V1 years in state 1 and Life B: V2 years in health state 2. . Thus, those respondents randomised to the 'reverse TTO' variant were presented with direct choices that used their earlier responses directly, with no need for any manipulations or assumptions to be made.

Analysis

Hypothesis One

If TTO responses are robust to the use of an iterative or non -iterative elicitation procedure- all else equal- then we would expect valuations in the non-iterative procedures to be the same as in their iterative equivalent as set out in Table 1. Formally this tests, for each health state valued: Ho: $U(X)_{Group\ 1} = U(X)_{Group5}$: Ho: $U(X)_{Group\ 2} = U(X)_{Group6}$, Ho: $U(X)_{Group3} = U(X)_{Group7}$, Ho: $U(X)_{Group4} = U(X)_{Group8}$

Hypothesis Two

If TTO responses are robust to valuing health states sequentially or concurrently, all else equal- then we would expect valuations in the 'concurrent' groups to be the same as in their 'sequential' equivalent. Formally this tests, for each health state valued:

Ho: $U(X)_{Group\ 1} = U(X)_{Group3}$: Ho: $U(X)_{Group\ 2} = U(X)_{Group4}$, Ho: $U(X)_{Group5} = U(X)_{Group7}$, Ho: $U(X)_{Group6} = U(X)_{Group8}$

These are tested using simple t tests of difference in means using a 5% level of significance.

Hypothesis Three

If responses to traditional 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. For each group we test this by a one- sample binomial test of whether the probability of choosing Life A was significantly different than 0.5. In addition, we 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 GEE model that allows for clustering of observations on individuals. We include the 'reverse TTO' group in order to test whether relaxing the assumptions of Utility Independence and CPTO have any influence on the main findings.

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 4 and compared to a representative UK population. It can be seen that there is some under-representation of young males and elderly females, a pattern that has been found previously in internet samples (Robinson, Gyrd-Hansen et al. 2013).

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 look at hypothesis one i.e. whether the TTO values themselves seem to be robust to the elicitation procedure used- iterative or non -iterative – holding all else equal. Comparing the TTO valuations of groups 1 and 5, no significant differences were found for 3 of the 4 health states valued[6] and no significant differences were found between groups 2 and 6 for any of the 4 states valued there[7] Likewise, there were no significant differences between valuations in groups 3 and 7 or 4 and 8. Hence, the impact of using an iterative or non-iterative procedure does not appear to have a significant impact on mean TTO valuations.

We next look at hypothesis two i.e. whether the TTO valuations were robust to whether states were valued ‘sequentially’ or ‘concurrently’ –as we argued that the latter will also make the TTO task less ‘transparent’ to respondents. Comparing the TTO valuations of groups 1 and 3, there was no significant differences in the case of 3 of the 4 states valued[8] and no significant differences detected between groups 5 and 7[9]. Likewise, there were no significant differences between groups 6 and 8 for any of the 4 heath states valued by those groups. In contrast, significant differences were found in mean TTO valuations between groups 2 and 4 for all 4 health states[10]. Hence, the evidence is somewhat mixed for impact of valuing health states concurrently as opposed to sequentially as in the traditional TTO procedure.

Overall, it seems that the TTO is reasonably robust to the procedural variations tested here. We now turn to hypothesis 3 and explore to what extent the aggregate choices coincide with TTO responses. 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. 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

[6] The p values for differences in mean valuations were: 11121 (p=0.145), 11121 (p=0.000), 21211 (p=0.498), 13122 (p=0.803).
[7] The p values for differences in mean valuations were: 13122 (0.703), 13224(0.293), 23242 (0.091),23314 (0.284)
[8] The p values for differences in mean valuations were: 11121 (p= .000)21211 (p=0.414), 12212 (p=0.761), 13122(p=0.236).
[9] The p values for differences in mean valuations were: 11121 (p=0.604), 21211 (p=0.575), 12212 (p=0.521), 1312 (p=0.976).
[10] The p values for differences in mean valuations were: 13122 (p= 0.005), 13224 (0.015),23242 (0.001), 23314 (0.009).

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 method	**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 (172) .000*	53:47 (168) .396	75:25 (165) .000*	52:58 (166) .103	72:28 (165) .000*	79:21 (163) .000*
Iterative: concurrent	Group 3	63:37 (179) .001*	64:36 (177) .000*	72:28 (170) .000*	62:38 (175) .290	73:27 (169) .000*	80:20 (168) .000*
Non iterative: sequential	Group 5	63:37 (140) .003*	55:45 (146) .282	69:31 (134) .000*	52:48 (143) .0198	68:32 (151) .000*	79:21 (133) .000*
Non-iterative: concurrent	Group 7	69:31 (143) .000*	62:38 (138) .005*	77:23 (138) .000*	54:46 (138) .162	72:28 (148) .000*	76:24 (145) .000*
Reverse TTO		79:21 (165) .000*	81:19 (157) .000*	87: 13 (150) .000*	74:26 (156) .000*	83:17 (149) .000*	84:16 (144) .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 method	**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 (152) .001*	68:32 (158) .000*	67:31 (152) .000*	61:39 (145) .319	52:48 (151) .625	43:57 (145) .135
Iterative: concurrent	Group 4	65:35 (141) .001*	64:36 (132) .001*	64:36 (135) .001*	57:43 (123) .471	73:27 (130) .000*	55:45 (118) .311
Non- iterative: sequential	Group 6	63:37 (120) .005*	61:39 (126) .016*	64:36 (118) .002*	50:50 (121) .585	45:55 (131) .294	53:47 (112) .637
Non-iterative: concurrent	Group 8	57:43 (106) .025*	59:41 (104) .062	65:35 (107) .000*	56:44 (107) .082	42:58 (112) .299	40:60 (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.

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 involving the more severe states –in the last 3 columns are closer to 50:50. We return to this issue in the discussion.

Table 3 shows the results of the 'reverse TTO' where the choices were set up in a way that did not depend on the assumptions of Utility Independence and CPTO holding. It can be seen that, rather than bring TTO and choice closer together, the reverse TTO has resulted in splits that were even further away from 50:50 and the null that p = 0.5 is rejected in every case. The reason is unclear, but the finding would appear to rule out failure of Utility Independence and CPTO being the main driver behind the results uncovered elsewhere.

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 it was the strong pattern to choose Life A that resulted in the move away from 50:50 in the splits, a positive co efficient here is effectively indicating choices more in line with 50:50- 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 subgroups of population. 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 'sequentially non-iterative' 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 any impact of TTO variant differed between 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
Reverse TTO	-0.8171	0.1377	-5.93	**0.000**	-1.0871 to -0.5471
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.*
*** There is no interaction term for the 'reverse TTO' as this was only carried out using the 'odd' states.*

Table 5 shows that, compared with the choices of the respondents in the traditional TTO – those completing the 'reverse TTO' are significantly different. But, rather than moving the choices closer to 50:50, the 'reverse TTO' has caused TTO and choice to diverge further in that the probability of choosing Life B is lower than in the traditional TTO (and it is the preference for Life A that is driving the disparity between TTO and choice). 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). This confirms the findings set out in Table 3 and again seems to rule out that the divergence from 50:50 in the traditional TTO is a failure of Utility Independence and CPTO assumed elsewhere.

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. Thirty six respondents valued 21211 more highly than 11121 in TTO and 89% of those then chose Life A in direct choice – involving more years in 11121. Whilst we could not know in advance how respondents would value each state in the TTO, the state that appeared in Life A was always the one that more people valued more highly than the other-than vice versa. Whilst this had clearly not been our intention, it has revealed an interesting issue in that the pattern of choices deviate from 50:50 in a systematic way.

We begin by looking at the cases of dominance as clear predictions may be made there (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. For the case of dominance in the odd groups, between 91% and 97% of those respondents who had valued the two states equally in TTO chose the dominant state in a straight choice. For the cases of dominance in the even groups, the numbers choosing the dominant state in the straight choice is less overwhelming, but the numbers valuing the two states equally in TTO are much smaller, making comparisons difficult. Respondents who valued the dominated state more highly in TTO would also be expected to choose Life A in the straight choice involving a longer time in a state that is strictly better, and this is supported, although the numbers are generally small.

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. More respondents in the odd groups valued both states equally in TTO, due in part to 'non- trading' behaviour and valuing both states at '1'. For example, 37 of the 76 respondents in group 1 who valued 11121 = 21211 in TTO were 'non-traders' for both states in the TTO. If respondents were simply reluctant to trade off life years, then states may be valued equally in TTO even although the respondent recognised that one was in fact better than the other. Given the way that we set up the direct choices here[11], they are not then required to trade off life years in the direct choice, a factor which may have shaped the results. This is less marked in the even groups where there were many fewer 'non traders' for both states. For example, only 4 of the 20 respondents in group 2 who valued 13122 = 13224 were 'non traders' in both states. We return to this issue in the discussion.

[11] In an actual DCE involving health states and duration, respondents would be asked to trade off life years in DCE too.

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	U1> U2	47% *(60)*	40% *(93)*	66%*(104)*	39% *(88)*	62% *(96)*	64% *(73)*
Iterative: sequential (traditional TTO)	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	U1> U2	39% *(83)*	49% *(112)*	62%*(120)*	41% *(93)*	64% *(108)*	65% *(80)*
Iterative: concurrent	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	U1> U2	42% *(62)*	52% *(90)*	63%*(104)*	34% *(67)*	60% *(98)*	66% *(67)*
Non-iterative: sequential	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 the procedure more in line with how choices would be presented in a DCE that set out to derive utility values using time as the numeriare-sometimes referred to as DCE_{TTO}. We found that TTO responses were fairly robust to the procedural variations tested here, which is to be welcomed particularly as previous studies have found that different procedures yield different results (Arnesen and Trommald 2005, Attema, Edelaar-Peeters et al. 2013). For example, it has previously been shown that the elicitation procedures used (Lenert and Alan 1998), whether 'props' are used or not (Dolan, Gudex et al. 1996) and the mode of administration of the survey (Norman, King et al. 2010) can all affect 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. We had not intentionally set the choices up in that way, but doing so has revealed an interesting finding. 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.

A priori, it seemed that an obvious reason that TTO and choices would not coincide was that the assumptions we used to generate the choices from TTO (utility independence and CPTO) do not reflect preferences well. Mutual Utility Independence has been shown to fail when health states present the "Maximum Endurable Time" effect, but we restricted our choices to states better than dead here, so this should reduce the impact of failures of MUI. The assumption of CPTO may be more problematic so we included the 'reverse TTO' in which neither MUI nor CPT are assumed. If anything, the reverse TTO resulted in choices that were further away from 50:50 than in the other variants, thereby ruling out failures of MUI and CPTO as the main driver of our results. Furthermore, the pattern of results uncovered cannot easily be explained by failures of MUI and CPTO in any case.

Certain 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 an overwhelming 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 easily be explained by insensitivity in the TTO and a reluctance to trade off life years. Given how we set up the choices here, such respondents were not asked to trade off life years in the direct choice, and they could then choose their 'preferred' health state without having to sacrifice life expectancy. This is, however, a function of how we chose to set up the direct choices here which would not happen in an actual DCE that allowed health states and duration to vary[12], so we must be cautious in generalizing here. Nevertheless, it does reveal an interesting pattern that helps shed light on the differences between TTO and choice. 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. Certain of this may be due to 'non-trading' in TTO for either health state and may be seen as another manifestation of insensitivity of TTO, particularly for the more mild states. But only around half of respondents who valued 11121 equal to 21211 in the TTO were non traders on both states, so there is something additional going on that may be explained 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'(Krantz 1967). Likewise, Tversky (1972) 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 may be happening in choices between health

[12] DCEs have been developed to estimate utility values that don't include duration, but they are not the focus of this paper.

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.

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

The study was funded by a grant awarded 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 views are the authors’ own and there are no conflicts of interest.

REFERENCES

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.

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.

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.

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.

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.

Delquie, P. (1993). "Inconsistent Trade-Offs between Attributes - New Evidence in Preference Assessment Biases." Management Science **39**(11): 1382-1395.

Dolan, P., C. Gudex, P. Kind and A. Williams (1996). "Valuing health states: a comparison of methods." J Health Econ **15**(2): 209-231.

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**: 1057-1075.

Krantz, D. H. (1967). "Extensive Measurement in Semiorders." Philosophy of Science **34**: 348-362.

Lenert, L.and G. Alan (1998). "The effect of search procedures on utility elicitations." Medical Decision Making **18**: 76-83.

Lichtens.S and P. Slovic (1971). "Reversals of Preference between Bids and Choices in Gambling Decisions." Journal of Experimental Psychology **89**(1): 46-57.

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.

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.

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)

Robinson, A., D. Gyrd-Hansen, P. Bacon, R. Baker, M. Pennington, C. Donaldson and Q. T. EuroVa (2013). "Estimating a WTP-based value of a QALY: the 'chained' approach." Soc Sci Med **92**: 92-104.

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.

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

Tversky, A., S. Sattath and P. Slovic (1988). "Contingent weighing in judgment and choice." Psychological Review **95**: 371-384.

Appendix 1: 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 Life A and Life B each time.

Appendix 2: *The 16 combinations of Lives A and B allocated randomly in the 'odd' group*

Life A	Life B- years in normal health			
20 years 11121	4	8	12	16
20 years 21211	4	8	12	16
20 years 12212	4	8	12	16
20 years 13122	4	8	12	16

Appendix 3: *Reverse order 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

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

LIFE B

5 YEARS WITH	
NO problems in walking about	
NO problems washing or dressing oneself	
NO problems doing usual activities	
NO pain or discomfort	
NOT anxious or depressed	
FOLLOWED BY DEATH	

Which would you prefer?

o Life A o Life B

Click NEXT to continue

Appendix 4: 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%