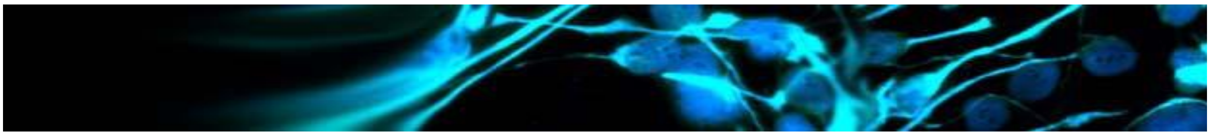


A comparison of alternative variants of the lead and lag time TTO

Office of Health Economics
Working Paper 10/02
24 December 2010

N Devlin[1], K Buckingham[2], K Shah[1], A Tsuchiya[3,4], C Tilling[3] , G Wilkinson[1], B van Hout[3]

1. *Office of Health Economics, London*
2. *Department of Preventive Medicine, University of Otago*
3. *School of Health and Related Research, University of Sheffield*
4. *Department of Economics, University of Sheffield*

Corresponding author:

[1] Professor Nancy J. Devlin, Director of Research, Office of Health Economics,12 Whitehall, London, SW1A 2DY. Phone: ++44 (0) 207 747 8858 Email: ndevlin@ohe.org

Key words: Time Trade Off, EQ-5D, duration, quality of life, QALYs

With some revisions, this paper has been published as: Devlin, N., Buckingham, K., Shah, K.K., Tsuchiya, A., Tilling, C., Wilkinson, G. and van Hout, B. (2012) A comparison of alternative variants of the lead and lag time TTO. Health Economics. Early view article. Available at: doi 10.1002/hec.2819

A comparison of alternative variants of the lead and lag time TTO

Abstract

Background: The estimation of Quality Adjusted Life Years gained from treatment requires length of life to be quality adjusted by the weight ('value') attached to the quality of life in each health state. These weights are anchored on a scale of 1 for full health and 0 for dead, with health states considered to be so bad that they are worse than being dead, having negative values. A widely used method for obtaining these values is the 'Time Trade Off' (TTO). The National Institute for Health and Clinical Excellence (NICE), for example, recommend the use of TTO values in evidence submitted to it on the cost effectiveness of new technologies. However, there are some important problems with TTO. These problems centre on the inability of the method adequately to handle very poor states of health, which people may consider to be worse than dead. Where that arises, the TTO has to switch to a different questioning process, with corresponding problems for the comparability and interpretation of values in the negative range. In previous research, we tested a new TTO approach, the 'Lead Time TTO', which is capable of producing weights both for states better and worse than dead in a uniform manner.

Aims: The aims of this research are (i) to investigate the values generated from Lead Time TTO (LT-TTO) using different combinations of the duration of the health state and the time in full health which participants are asked to consider; as well as varying the order in which these appear (Lag Time TTO); (ii) to gauge if values generated from these methods concur with participants' views as to whether the states are better or worse than dead (iii) to explore a range of methods for handling the preferences of those whose distaste for very poor health states is such that they 'use up' all their lead time.

Methods: A sample of 200 members of the general public valued five health states, using two of four variants of the LT-TTO: a lead time of 10 years with a health state duration of 20 years; a lead time of 5 years and a health state duration of 1 year; a lead time of 5 years and a duration of 10 years; and a duration of 5 years with a lag time of 10 years. Participants also responded to a range of supplementary tasks and other questions.

Results: Values are influenced by the length of the lead time relative to the health state duration. Longer lead times enable somewhat more preferences to be captured, but appear to exert a framing effect on values. Lag time TTO results in less non-trading for mild states, and to participants trading off less time for severe states. Of those who valued the worst health state as negative, 70% also expressed the view that this state was worse than dead.

Conclusions: LT-TTO confers an important advantage over the traditional TTO by providing a single method capable of generating positive and negative values that seem broadly in keeping with participants' stated views about those states being better or worse than dead. However, values are sensitive to the length of time in full health relative to the duration of the state to be valued, and to the order in which these appear (lead vs. lag time). For those who use up their lead time, we show that additional ways of eliciting these preferences (via additional questioning) are feasible, as is modelling those values (via survival analysis). However, a small (<5%) group of participants remain whose preferences are so 'extreme' they cannot be captured by any approach.

1. Introduction

Time Trade Off (TTO) is one of a number of a stated preference methods developed for valuing health states, yielding interval scale values anchored at 1 (full health) and 0 (dead) as required in the calculation of Quality Adjusted Life Years for cost effectiveness analysis. The TTO is one of the most widely used methods for that purpose, and most of the national 'value sets' for EQ-5D have used the TTO. For example, the EQ-5D value set routinely used by NICE in health technology appraisal (NICE 2008) is based on TTO values elicited from the UK general public in the MVH study (Dolan 1997), as are the EQ-5D values sets for the US (Shaw et al 2005), Japan (Tsuchiya et al 2002), the Netherlands (Lamers et al 2006) and France (Chevalier and de Pouvourville 2009) (for an overview, see Szende et al 2007).

TTO establishes the value for a health state by finding the amount of time in full health (x) which is considered equal in utility terms to a given amount of time in a poor health state (t), and calculating the utility 'flow' of the state[1] as (x/t). In EQ-5D valuation studies t is set (by convention) at 10 years.

Despite its widespread use, there are acknowledged problems with TTO – particularly in relation to how to value health states that are considered 'worse than being dead' (Tilling et al 2010). The TTO task, described above, does not work for the elicitation of preferences for states worse than dead where, by definition, no amount of time in good health can be equivalent to an amount of time in a state worse than dead. Obtaining those values therefore requires the introduction of a different preference elicitation task.

To value a state worse than dead, conventionally, the participant is asked to choose between immediate death, and spending a length of time (t-x) in the health state of interest, followed by x years in full health. x is varied until the participant is indifferent between the two options. The value of state H_i is given by $U(H_i) = -x/(t-x)$.

There are four important, related problems with this approach:

(a) States better than dead (SBD) and worse than dead (SWD) are valued using a completely different elicitation procedure – so the values are non-comparable. The *aggregation* of positive and negative values, as required in the calculation of mean values and estimation of value sets, is of questionable validity.
(b) Whereas TTO values for SBD are obtained by varying x while t is fixed, the valuation of SWD involves *simultaneously* changing both the numerator and the denominator, to add up to a fixed number (conventionally, 10 years)[2]. In the absence of clear evidence about the role of constant

1 The phrase 'value flow' is usually rendered simply as 'value', a convention that we shall generally maintain throughout the paper.
2 In valuing a SBD, a move from 2 years to 1 year in full health reduces the estimate of utility from 0.2 to 0.1, a change of 0.1; trading one year in full health will always result in a 0.1 change in utility. In valuing a SWD, a move from 2 years to 1 year in full health decreases the estimate of utility from -4 to -9; a change of 5. Furthermore, the size of this decrement will depend where on the scale the choice is being made. If it is a move from 8 to 7 years, utility decreases from -0.25 to -0.43, a change of 0.18. The key issue is that the thing being valued, time in poor health, is changing.

proportionality on TTO values, arguably all health states should be elicited for specified durations – a condition which is not met if we vary duration in poor health (Buckingham and Devlin 2009).
(c) The minimum possible value is driven by a researcher decision about the smallest unit of time that can be used in the task[3].
(d) The approach produces extreme negative values and these cast doubt on whether these have any real meaning – so it has become standard practice to perform various *ex post* transformations to bound negative valuations to -1 (Lamers 1997) or to propose other ways of manipulating these data (Craig and Oppe 2010).

These issues have serious implications for the use of TTO values in HTA, where the estimation of QALYs requires health state values to be measured on the same scale i.e. a change from -0.5 to 0 is assumed to mean the same thing as a change from 0 to 0.5.

An alternative approach to TTO, identified by Robinson and Spencer (2006), provides a potential solution to these problems as it comprises a uniform elicitation procedure capable of yielding both positive and negative values. The method adds extra time in full health before each of the alternatives usually presented in the TTO which makes it possible for participants to express their distaste for very poor health states by 'trading into' these additional years of full health. In Figure 1a, 'option A' is spent in full health (followed by death), whereas 'option B' involves 10 years in full health, then 5 years in the health state in question (H_i) (followed by death). The objective of the exercise is to identify the point where the respondent is indifferent between the two options, by changing the timing of death in option A. When the health state is better than being dead (SBD), then example (i) in Figure 1a applies: at the point of indifference, the duration in full health in Option A will be longer than the duration in full health in Option B. When the health state is worse than being dead (SWD), then example (ii) applies: the duration of full health in Option A is *less* than the duration of full health in Option B.

We have shown that this 'lead time' TTO (LT-TTO) is feasible for the valuation of EQ-5D states (Devlin et al 2010). However, a number of questions were identified as requiring further research in order better to understand the potential of LT-TTO as a valuation method.

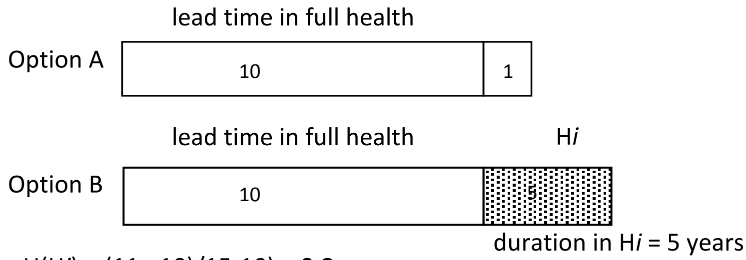
First, there is no theoretical basis to guide the selection of the length of the lead time to accompany any given duration. What ratio of lead time to health state duration is required adequately to capture preferences regarding very poor health states is unknown, as is the way that the various combinations of duration and lead times affect the valuation data more generally.

Second, the lead time approach has a logical analog: the additional time in full health could be placed *after* the health state being valued, rather than before it (Verschuuren 2006) to form a 'lag time' TTO – see Figure 1b. As can be seen, the difference between this and Figure 1a is simply that in Option B, the ordering of full health and H_i is reversed: (iii) illustrates this for SBD and (iv) for SWD. Note that while the ordering in Option B can be reversed, the same cannot be done for Option A, since this would involve being dead before being in full health. Given that the lag time TTO is a plausible alternative to the lead time TTO, research is required to identify the effect of the temporal placement of the additional time in full health. In the remainder of this paper, we use the term LT-TTO to include both

[3] For example, if the minimum trade is 6 months, the minimum value is -19; if it is 3 months, then -39.

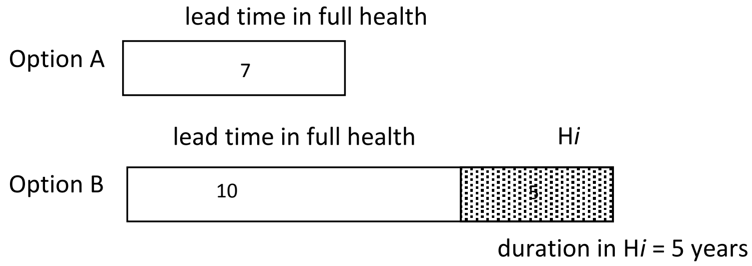
Figure 1a. An illustration of the lead (i and ii) time approach, with lead time = 10 years and duration = 5 years and (ratio 2:1), for positive and negative values respectively.

i. Lead time: at the point of indifference, the state happens to be 'better than dead'



U(H*i*) = (11 - 10)/15-10) = **0.2**

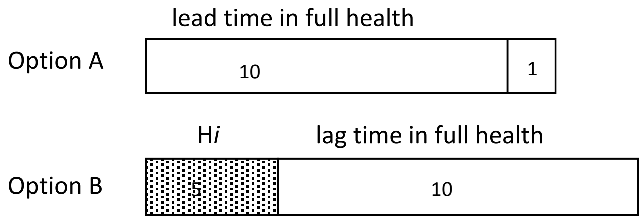
ii. Lead time: at the point of indifference, the state happens to be 'worse than dead'



U(Hi) = (*7* - 10)/15-10) = **-0.6**

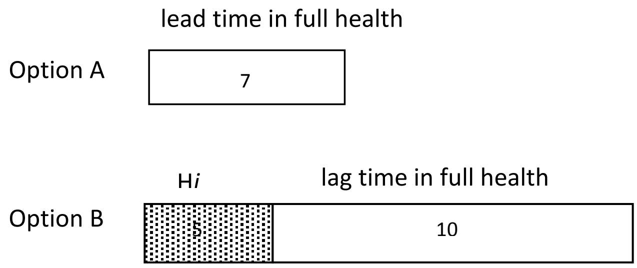
Figure 1b. An illustration of the lag (iii and iv) time approaches, with lag time = 10 years and duration = 5 years and (ratio 2:1) for positive and negative values respectively.

iii. Lag time: at the point of indifference, the state happens to be 'better than dead'



U(H*i*) = (11 - 10)/15-10) = **0.2**

iv. Lag time: at the point of indifference, the state happens to be 'worse than dead'



U(Hi) = (*7* - 10)/15-10) = **-0.6**

lead and lag time variants.

Third, Devlin et al (2010) provide tentative evidence that the lead time approach might affect the distribution of values not just for SWD, but also for SBD. More needs to be known about the way the LT-TTO might affect mean and median values across the EQ-5D descriptive system.

Fourth, a key feature of the LT-TTO is that the iterative trading process can allow a participant to move between negative and positive valuations without explicitly considering whether a state is worse than or better than being dead. This has the important merit of avoiding the 'focusing effect' that might accompany such a discussion. However, it raises the question of whether the resulting negative values concur with participants' stated views about those states as being worse than dead.

Finally, when participants exhaust their lead time, this indicates they would have liked to give a value less than the minimum value allowed by the design. As such, the valuation data are "censored" - suggesting that

survival analysis may provide a means of modeling the distribution of these unobserved data. To our knowledge, survival analysis has not previously been applied to the analysis of stated preferences data.

The aim of this study is to provide evidence on these issues, and to facilitate a better understanding of the way that the LT-TTO influences participants' valuations of health states.

2. Research design

Our study proceeded in two phases. In phase 1, a small pilot study was conducted to test 5 contrasting variants of the LT-TTO and to inform the selection of the variants to include in phase 2; to test the feasibility of a prototype digital aid; and to guide the development of structured feedback questions.

Phase 1: Pilot study

Four variants of the lead time, and one lag-time, TTO were selected to conduct initial tests of feasibility – see Table 1. Each used a *higher* ratio (of lead time to duration) than used in previous research (1.5:1), as our earlier results indicated more lead time is required, relative to the duration of the state, to allow participants to express their distaste for severe health states. Variant [i] uses the 'standard' duration of 10 years, but combines that with a longer lead time (20 years) than included in our previous work (ratio 2:1). However, longer lead times added to the conventional duration of 10 years might quickly run into issues of plausibility with older participants. Variant [ii] halved the duration previously used in TTO studies (to 5), and combined it with a lead-to-duration ratio of 3:1. The overall 'profile', 20 years, is shorter than that used in our feasibility study, but may still present issues for elderly participants. We therefore also included much shorter time profiles, although these may be less credible to younger respondents. Variant [iii] uses a duration of just one year – which allows us to experiment with a further increase in the lead-to-duration ratio of 5:1. Variant [iv] is the corresponding 'lag time' presentation: it uses the same duration and ratio as [iii], but alters the temporal placement of the additional 'trading time' in full health. Finally, variant [v] also used a one year duration, combining that with an 'extreme' ratio (10:1). This was intended to allow us to explore the possibility that, even when this ratio is very high, there would remain a small proportion of participant who will exhaust lead time.

Table 1. LT-TTO variants included in phase 1 (pilot study)

	[i]	[ii]	[iii]	[iv]	[v]
Duration (years):	10	5	1	1	1
Lead time (years):	20	15	5	n/a	10
Lag (years)	n/a	n/a	n/a	5	n/a
Ratio† of lead(lag):duration	2:1	3:1	5:1	5:1	10:1

†*Note*: Our previous research used ratios of 1:1 (10,10) (Devlin et al 2009) and 1.5:1 (15,10) (Devlin et al 2010).

Interview scripts and a digital aid *('Time Trader'*[4]) were developed to accommodate each of these variants. The digital aid mimics the presentation of the TTO used in the MVH protocol e.g. Option A and Option B are displayed horizontally and using the same colours to represent each state – see Figure 2. In addition to acting as a visual 'prompt', *Time Trader*, directly captures all participants' responses to the tasks, and records the time taken to complete them. It also has functionality to automate the iterative process for the trades offered to participants. However, for the pilot study we provided a manual over-ride, to allow interviewers to explore participants' responses in more detail, and to inform the fully automated process to be used in the main study.

Each participant was asked to value 3 states (33333, 22222, 23323: the most severe state, a moderate state and one intermediate to those), using 3 of the design variants in Table 1 (= 9 TTO tasks per participant). The variants were presented in 5 different combinations and orders (a 'Group') with no fewer than 5 participants in any given 'Group'.

The sample comprised undergraduate students in London and Sheffield (n=20) and a small sample of people over 60 years (n=5) in Aberdeen. Interviews were undertaken by KS, CT and KB, and on average took 36 minutes (s.d 17).

Results suggested *all* of the variants in Table 1 were feasible for the participants, but variants with high lead-to-duration ratios posed a challenge in terms of allowing sufficient 'granularity' in the trade-offs for SBD. For example, variant [v] can capture values as low as -10, but the implication of this is that the health state in question only takes up one 11th of the length of Option B. As a result, the length of visual 'space' provided on the computer screen for trade-offs for SBD in this variant is limited. This problem is likely to be particularly important for mild states (where trading occurs at the right hand end of the scale in Option B) and potentially affects not just the values of those who 'trade', but also the extent to which non-trading behaviour may be observed.

[4] Available from the authors on request.

This suggested that the choice of duration and lead time variants needs to address an apparent trade-off between: (i) having a (sufficiently high) level of lead time to reduce the exhaustion of lead time for SWD to an 'acceptable' level, and (ii) having a (sufficiently low) level of lead time, to ensure an 'acceptable' level of acuity.

Phase 2: Main study

In the light of results from the pilot study, variants for the main study were chosen to focus on those with somewhat lower lead-to-duration ratios – see Table 2.

Table 2. LT-TTO variants selected for the main study of the general public

	[a]	[b]	[c]	[d]
Duration (years):	10	1	5	5
Lead time (years):	20	5	10	n/a
Lag (years)	n/a	n/a	n/a	10
Ratio* of lead(lag):duration	2:1	5:1	2:1	2:1

Variant [a] was selected as it uses the same duration conventionally used in TTO valuation of EQ-5D states. Although we had been concerned about the potential feasibility problems associated with the length of the overall profile, the pilot study did not provide evidence of that – and inclusion of a longer overall profile is useful for comparative purposes. Variant [b] allows us to explore LT-TTO values for a 1 year duration; this is combined with a lead-to-duration ratio of 5:1, to permit inclusion of a high lead:duration ratio, while reducing the concerns associated with variant [v] in the pilot. Finally, Variants [c] and [d] each use durations of 5 years (these are the variants illustrated in Figures 1a and b earlier).

We considered, and rejected, pairing the 5 year duration with a 5 year lead/lag time because previous research suggested that a 1:1 ratio is insufficient for capturing a large proportion of values < -1 for severe states. Also, reducing the ratio to 1:1 reduces the contrast between the lead and lag time variants – and comparing these approaches, and understanding the differences in the valuation data they produce, is a key part of the study. Therefore, the 5 year duration is paired with a 10 year lead time.

Each participant received 2 of these key variants of the LT-TTO task, to value 5 states. Four sub-groups of 50 participants each valued the pairs of designs, as shown in Table 3. This design means that every lead time variant is paired with every other lead time variant, but the lag time variant is only paired with its corresponding lead time variant. In total, 2 sub-groups of 50 (n=100) were to complete variant [a]; 2 sub-groups of 50 (n=100) to complete variant [b]; 3 sets of 50 (n=150) to complete design [c], and n=50 will complete design [d].

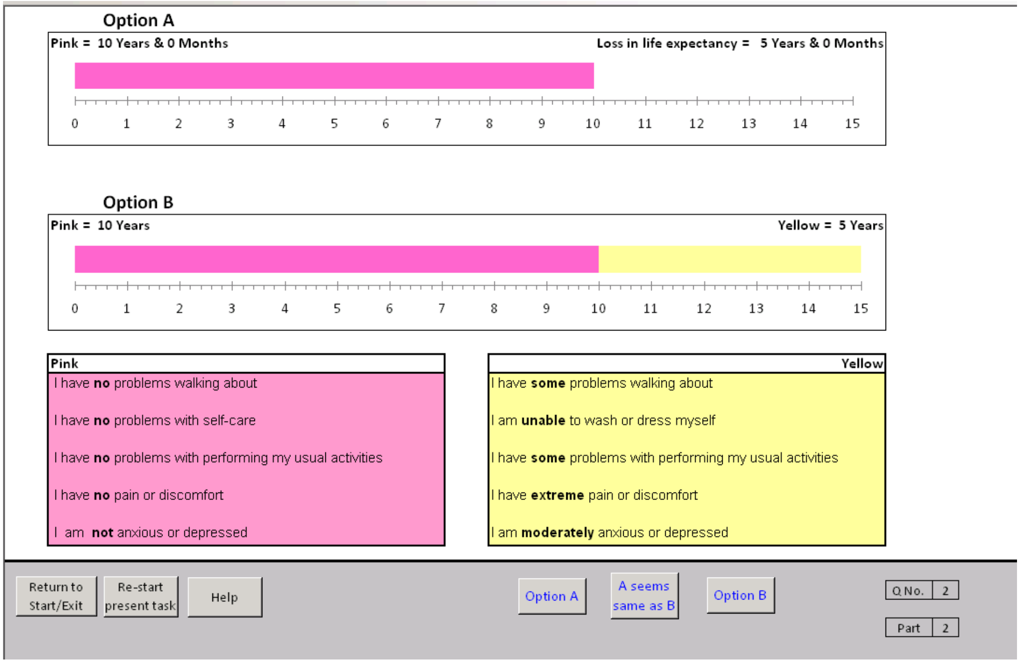
Table 3. Participant sub-groups in Phase 2 and the variants to be used in each

Group	Variant pairs
1	[a] + [b]
2	[a] + [c]
3	[b] + [c]
4	[c] + [d]

All LT-TTO tasks were administered using *Time Trader* (see Figure 2) to display tradeoffs and record responses. Other data were sought and recorded in a participant booklet. This included, before the LT-TTO tasks, self-rating of health on the EQ-5D; ranking of a set of 7 health state cards (the 5 health states to be valued plus 11111 and dead) and, following the LT-TTO tasks, background characteristics and structured participant feedback questions. Where participants exhausted their lead time in valuing a state, they were also asked a series of follow-up questions in an attempt to locate their point of indifference by (a) keeping the duration of the health state unchanged, but sequentially extending the amount of lead time and (b) keeping the lead time the same, but sequentially decreasing the duration of the poor health state.

Participants' feedback from the pilot study was used to further improve the digital aid. For example, on-screen information was added to display the number of years in Options A and B, in addition to the time scales shown. The visual display of time in state 'dead' in Option A was dropped as being superfluous. The iteration process was fully automated. Each task started with the length of full health in Option A set equal to the length of full health in Option B (as shown in Figure 2). The trades then proceeded using a 'bi-section' approach (adding or subtracting half the time in Option A), modified somewhat to select 'natural' periods of time. Further, a 'magnification' procedure was developed for use in the main study, to increase acuity for the valuation of SBD. This allowed the interviewer the option of 'zooming in' on just the right-hand section of the scales that are relevant to positive values.

Figure 2. Screen shot of *Time Trader*, illustrating the use of variant (c) to value state 23232



Selection of states

EQ-5D states were selected from those included in the MVH (see Table 4). State 33333 was included in order to gauge the adequacy of the variants in capturing disutility for the most severe state. However, given the special status of 33333 as the recognisably worst state, we also included one other severe state (23232). Each participant valued a set of five states using two LT-TTO variants. In order to expand somewhat the number of mild states that could be included in the study, we assigned each participant to one of two sets of states, increasing the number of states for which we seek values to 7. We considered but rejected the randomisation of states in the interviews: state 33333 needed to be the last state valued in each interview, as this was accompanied by a question to seek the participant's opinion about whether that state was better or worse than being dead. The other states were presented in a fixed 'pseudo random' order in each LT-TTO task in each variant: the relatively small size of the sample sub-groups generated by the complex study design and our wish to compare values across sub-groups meant we wanted to ensure states were valued in the same order across variants and respondents.

Table 4. EQ-5D states for which values were sought (with MVH means and medians†)

Set 1		Set 2	
23232	(-0.10; -0.08)	23232	(-0.10; -0.08)
33333	(-0.54; -0.65)	33333	(-0.54; -0.65)
11112	(0.82; 0.93)	11112	(0.82; 0.93)
11122*	(0.72; 0.83)	12111*	(0.83: 0.93)
11211*	(0.87; 0.95)	22121*	(0.64; 0.78)

*States which appear in one set and not the other.

† MVH means and medians with transformation to negative value -1 (source MVH Group, 1994)

Supplementary tasks for participants who use up their lead time

The LT-TTO variants included in this study increase the amount of lead time relative to duration, compared with earlier studies, in an attempt to capture a greater proportion of negative values. However, it is possible that some participants may still use up all their lead time in valuing the very worst states. We included two supplementary tasks for these participants who exhaust lead time, both aimed at locating the participant's point of indifference: keeping the duration of the poor health state the same, but extending the lead time and keeping the lead time the same, but reducing the duration. If a respondent does not reach a point of indifference after two supplementary trade off questions, this is regarded as 'exhausting supplement'.

In addition, the extending lead time question had an open question at the end, asking "How long a time in the pink health state would you require in order for you to choose Option B over immediate death?" If a respondent answers that no length of time would make them choose Option B, then, taken at face value, this implies that the health state has an infinitely low value.

The options which were offered are described on Table 5. These supplementary questions were asked and answers recorded by the interviewer rather than by *Time Trader*. Where a participant switched their preference from Option A to Option B, the value was taken as the mid-point between the most two recent options presented.

Table 5. Supplementary trade-off tasks used to establish indifference for those who used up their lead time

Variant	Ratio of lead(lag):duration	Extended leads		Reduced durations	
		Iteration 1	Iteration 2	Iteration 1	Iteration 2
a	20:10	30:10	>30:10*	20:5	20:1
b	5:1	10:1	>10:1*	5:0.5	5:0.25
c	10:5	25:5	>25:5*	10:2	10:1
d	10:5	25:5	>25:5*	10:2	10:1

* Open-ended question asked of participants who preferred immediate death to Option B even with an extension to lead time: "How long a time in the pink health state would you require in order for you to choose Option B over immediate death?"

3. Data collection

Interviews were conducted by a team of professional interviewers employed by Oxford Outcomes, and took place during April and May 2010. The interviewers all had prior experience in conducting TTO; additional training was provided to introduce them to the LT-TTO and to rehearse use of the interview scripts, participant booklets and Time Trader.

The sample comprised a panel of members of the general public who had participated in previous studies with Oxford Outcomes and had indicated their willingness to be contacted again. None of the participants had previously taken part in an EQ-5D valuation study. The sample was recruited to be broadly representative of the general population with respect to age and sex. Participants were paid £25 for their involvement.

4. Methods of analyses

LT-TTO data captured by the digital aid was linked with data entered from the participant self-completion booklets via participants' ID number.

The distribution of LT-TTO values by state and by variant were compared by frequency distributions, and by categorising the respondents into non-traders (with a value of 1); those with positive values; those with negative values within the minimum possible using the lead (lag) time; those with negative values using either the tasks; and those with negative values who could not achieve a point of indifference under any task (ie. exhausted the supplementary tasks). Comparisons of values from the lead and lag time variants ([c] and [d]) included testing for potential ordering effects (whether [c] preceded [d] or vice versa).

Means, medians and variance in values by state and by variant were calculated under each of a range of treatments of the participants who exhausted their lead time e.g. excluding those responses; including the additional values obtained by the supplementary tasks; and assigning remaining missing responses that the exhaust supplementary tasks an arbitrary value of -100. The distributions and cumulative frequencies of positive and negative values were examined to identify any differences between variants.

Probit analysis was used to predict the probability of non-trading behavior; and exhausting lead time; in each of the variants, with age and sex as potential confounders.

The data on the number of iterations required to establish indifference, which were captured by *Time Trader*, were used to compare averages between variants with respect to positive and negative values. A regression analysis was undertaken to examine the participant and valuation task characteristics that were associated with a larger number of steps to indifference.

The level of agreement between each participant's ranking of health states implicit in each their LT-TTO values and their direct ranking of those same states. This was examined by estimating Spearman's rho for each individual respondent. The level of agreement *overall* for each variant was taken as the average of the relevant sample's Spearman's rho.

Relationships between the participants' characteristics (for example, age, and experience of ill health in self and in others) and responses to the qualitative feedback questions, and between those feedback questions and values, were examined using Pearson's chi-squared.

When trading techniques are used to estimate values, those values are inferred at points of indifference. For example, when individuals are indifferent between 5 years in perfect health and a combination of 10 years in perfect health followed by 5 years in EQ-5D state 33333, then the implied value (flow) for 33333 is -1. However, if a respondent exhausts their lead/lag time without arriving at a point of indifference, a value cannot be obtained: they prefer dead to the combination of time in perfect health and the duration in the imperfect state. When the ratio between lead time and the duration of poor health is 2:1 and subjects prefer dead to a state such as 33333, they implicitly state that the value of this state is < -2. Their 'true' value is unknown – it might be -5 or -100. When the ratio between lead time and imperfect time is 5:1 and subjects prefer dead to for example 33333 they implicitly state that the value of this state is < -5. As a consequence, one may not be able to calculate an average value for the health states, as this average may be quite different depending on the 'true' negative values of those who preferred immediate dead. As such those values can be considered as *censored* observations, and the application of survival analysis is potentially relevant, where the problem noted above (estimating an expected survival while not everybody has died) is quite common. This can be done non-parametrically, using Kaplan Meier (or product limit estimates) and Cox regression analysis, or parametrically, assuming underlying distributions such as Weibull curves. We attempt an exploratory analysis using these techniques, and report selected results.

Statistical analyses were undertaken in STATA version 10, apart from survival analyses which were carried out in R.

5. Results

Interviews were conducted with 208 members of the general public. One interview was terminated due to the participant having difficulty following the interview questions; data are therefore available for 207 participants. Table 6 shows the background characteristics of the sample, with the corresponding MVH study results (where available) for comparative purposes. Compared with the MVH sample, our sample contained proportionally fewer homeowners. Like the MVH sample, our sample contained more females than males. A smaller proportion of our sample had experience of serious illness both in themselves and in their family, but a slightly higher proportion had experienced serious illness in caring for others. Mean self-reported health on the EQ-VAS was slightly lower amongst our sample.

Interviews were conducted by seven interviewers. There was considerable variation in sample composition across interviewers; for example, mean self-reported health on the EQ-VAS ranging from 0.77 to 0.89. The mean interview duration was 24 minutes overall; this ranged across the interviewers, with a minimum average interview length per interviewer of 18 minutes and a highest average of 34 minutes.

Table 6. Sample background and health characteristics

Variable		LT-TTO sample	MVH general public sample†
Gender	Male	43%	46%
	Female	57%	54%
Age: mean (sd)		37 (14.8)	39††
Number of people in household: mean		3.6	
Number of children in household: mean		0.3	
Employment status	In employment or self-employment	60%	59%
	Not employed	40%	41%
Education after minimum school leaving age	Yes	85%	
	No	15%	
Degree or equivalent professional qualification	Yes	59%	
	No	41%	
Home ownership status	Own home outright, or with a mortgage	51%	70%
	Does not own home	49%	30%
EQ-VAS: mean		0.83	0.86
Experience of serious illness in you yourself	Yes	19%	32%
	No	81%	68%
Experience of serious illness in your family	Yes	67%	72%
	No	33%	28%
Experience of serious illness in caring for others	Yes	31%	28%
	No	69%	72%

† Source: Kind et al. (1999)

†† Source: ONS (2009) - http://www.statistics.gov.uk/cci/nugget.asp?id=6 (accessed 14.06.2010)

Table 7 shows the categorised distribution of valuations by variant and by state. Negative valuations are further broken down into those that reached an indifference point using lead time, those that reached an indifference point by the supplementary exercises and those that exhausted the supplements and did not reach an indifference point under any of these procedures. The proportion of positive values is greater for the milder states (those with more dimensions on level 1 and no dimensions on level 3) as expected. There is some variation in this proportion by variant. None of the variants appears to have either a systematically higher or lower proportion of positive values across *all* states. However, variant [d] - the lag time TTO - has a markedly higher proportion of positive values for the two severe states, 23232 and 33333.

Table 7. Distribution of valuations by variant and state

			--------------------------------Negative values -----------------------------------			
Health state	**Variant**	**Positive values**	**Valued using lead time**	**Valued by extension and/or reduction**	**Not possible to achieve indifference**	**Missing**
	a	94%	4%	1%	0%	1%
	b	92%	8%	0%	0%	0%
11112	c	95%	3%	1%	1%	1%
	d	98%	2%	0%	0%	0%
	All variants	**94%**	**4%**	**0%**	**0%**	**0%**
	a	98%	2%	0%	0%	0%
	b	92%	8%	0%	0%	0%
11211	c	98%	2%	0%	0%	0%
	d	97%	3%	0%	0%	0%
	All variants	**96%**	**4%**	**0%**	**0%**	**0%**
	a	96%	2%	2%	0%	0%
	b	96%	4%	0%	0%	0%
12111	c	97%	1%	1%	0%	0%
	d	96%	4%	0%	0%	0%
	All variants	**96%**	**3%**	**1%**	**0%**	**0%**
	a	94%	4%	0%	0%	2%
	b	87%	13%	0%	0%	0%
11122	c	92%	6%	1%	1%	0%
	d	85%	15%	0%	0%	0%
	All variants	**90%**	**8%**	**0%**	**0%**	**0%**
	a	89%	9%	2%	0%	0%
	b	88%	13%	0%	0%	0%
22121	c	92%	7%	1%	0%	0%
	d	96%	4%	0%	0%	0%
	All variants	**91%**	**8%**	**1%**	**0%**	**0%**
	a	32%	63%	1%	3%	1%
	b	33%	62%	5%	0%	0%
23232	c	32%	61%	4%	3%	0%
	d	50%	41%	3%	3%	2%
	All variants	**35%**	**59%**	**4%**	**2%**	**0%**
	a	11%	65%	15%	3%	6%
	b	11%	67%	14%	4%	4%
33333	c	14%	67%	12%	5%	2%
	d	28%	45%	16%	10%	2%
	All variants	**14%**	**63%**	**14%**	**5%**	**3%**
All states	**All variants**	**66%**	**28%**	**4%**	**2%**	**1%**

There were 111 instances (out of 207 participants x 10 TTO tasks = 2,070 valuation tasks) of lead/lag time being exhausted. Participants who exhausted their lead time were asked a series of supplementary questions which sought to determine whether indifference could be achieved either by further extending the amount of lead (or lag) time; or by reducing the duration of the poor health state. A sizeable minority of respondents (3-5% in lead time variants [a] – [c], and 10% in the lag time [d]) were still unable to reach a point of indifference, in effect claiming the health state was so undesirable that no amount of lead time would be sufficiently long, and no duration in that health state would be sufficiently short, to compel them to choose Option B over immediate death.

Participants aged 60 or older were statistically significantly more likely to exhaust lead time than younger participants ($p < 0.001$). Gender was not found to affect the likelihood of lead time exhaustion.

Table 8. Means, medians and standard deviations of values by variant and by state

			------------Baseline results††	----------		-----------------Sensitivity analysis†††	-----------------		
Health state	**Variant**	**n**	**Mean**	**Median**	**sd**	**LTE**	**LTE2**	**HSR**	**HSR2**
	a	97	0.77	0.90	0.38	0.77	0.77	0.56	0.56
	b	100	0.57	0.88	0.86	0.57	0.57	0.57	0.57
11112	c	154	0.77	0.90	0.31	0.12	0.77	0.09	0.73
	d	58	0.81	0.90	0.23	0.81	0.81	0.81	0.81
	a	52	0.87	0.98	0.22	0.87	0.87	0.87	0.87
	b	52	0.63	0.96	0.78	0.63	0.63	0.63	0.63
11211	c	86	0.80	0.94	0.34	0.80	0.80	0.80	0.80
	d	34	0.80	0.90	0.33	0.80	0.80	0.80	0.80
	a	46	0.79	0.90	0.28	0.79	0.79	0.35	0.35
	b	48	0.71	0.84	0.35	0.71	0.71	0.71	0.71
12111	c	70	0.77	0.90	0.29	0.77	0.77	0.69	0.69
	d	24	0.76	0.81	0.35	0.76	0.76	0.76	0.76
	a	51	0.67	0.71	0.35	0.67	0.67	0.67	0.67
	b	52	0.36	0.74	1.09	0.36	0.36	0.36	0.36
11122	c	84	0.66	0.78	0.34	-0.56	0.61	-1.68	0.66
	d	34	0.56	0.66	0.49	0.56	0.56	0.56	0.56
	a	46	0.52	0.65	0.51	0.52	0.52	0.08	0.08
	b	48	0.47	0.77	0.72	0.47	0.47	0.47	0.47
22121	c	70	0.63	0.74	0.39	0.63	0.63	-0.79	0.63
	d	24	0.68	0.81	0.37	0.68	0.68	0.68	0.68
	a	95	-0.43	-0.43	0.73	-3.49	-0.41	-3.48	-0.44
	b	95	-1.02	-0.69	1.53	-1.35	-1.35	-3.35	-1.37
23232	c	146	-0.30	-0.40	0.71	-3.09	-0.54	-4.86	-0.42
	d	53	-0.08	0.00	0.78	-5.51	-0.26	-3.70	-0.20
	a	78	-0.70	-0.61	0.69	-10.58	-1.00	-7.32	-2.06
	b	80	-1.77	-1.50	1.62	-7.92	-3.92	-12.56	-2.40
33333	c	129	-0.68	-0.80	0.66	-6.34	-1.18	-8.77	-1.06
	d	42	-0.30	-0.39	0.85	-16.71	-1.10	-14.76	-0.84

†† Baseline means and standard deviations calculated using all values except for those associated with participants who exhausted all of the available lead time.

††† Sensitivity analysis results reported in terms of mean values under the following conditions: LTE = includes valuations obtained by extending lead/lag time; attributing a default minimum value of -100 to observations where indifference was not achieved using this method. LTE2 = includes valuations obtained by extending lead/lag time; excludes observations where indifference was not achieved using this method. HSR = includes valuations obtained by reducing health state duration; attributing a default minimum value of -100 to observations where indifference was not achieved using this method. HSR2 =

includes valuations obtained by reducing health state duration; excludes observations where indifference was not achieved using this method.

Note: Shaded cells indicate the variant with lowest mean (and median) value for each health state.

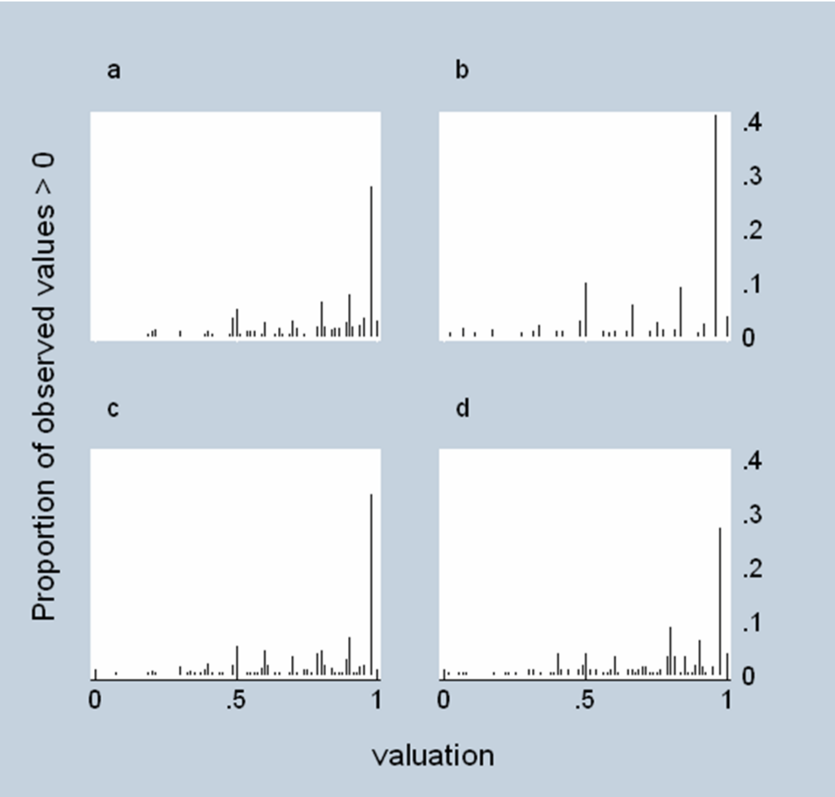
Table 8 summarises the values by variant and by state. Variant [b] has notably lower mean values than the other variants – it has the lowest means for *all* states in the base case (see shaded cells). This is not the case with the median values – however, variant [b] does account for the lowest base-case medians for the two most severe states. Variant [b] also has the largest standard deviation throughout, reflecting the wider range of values it can produce. Once the constraints of the minimum value possible within each LT-TTO variant are relaxed, by including the data from the supplementary tasks, the differences between means and medians produced by the variants become more mixed. Compared with the MVH results for the same health states, in general the mean and medians observed here tend to be somewhat lower, while the standard deviations are smaller (where negative values in the MVH data are transformed to -1; MVH Group, 1994 - see Table 4 above).

Fifty two participants were assigned to the matching lead and lag time variants ([c] and [d], respectively) – half started with the lead time variant; the other half with the lag time variant. The order in which lead and lag time appeared in the interviews had no effect on valuations (p=0.74).

Our pilot study had suggested a concern that variants with short health state durations might give rise to problems relating to visual 'acuity' because of the limited size of the computer screens, and that this would result in participants providing responses that clustered on 'round' values such as 0.5 and 1 (whereas in variants with longer health state durations, the larger visual 'space' may facilitate participants trading in finer units of time). The frequency distributions of values greater than zero are shown in Figure 3. At face value, these appear broadly similar and in all variants the mode is just below 1, representing a cluster of values for the very mild states that are just 'one step away' (given the iterative process) from full health. Variants [a], [c] and [d] share a mode of 0.975, with 31, 68 and 22 responses respectively. The mode in variant [b] is 0.96, with 43 participants sharing that value[5]. The histograms in Figure 3 suggest that the data generated by variant [b] has fewer unique positive values with responses centring on a smaller number of values.

[5] Recall that the study design required a larger number of variant [c] questions. As *proportions* of the values > 0, these modal values accounted for 33%, 47%, 46% and 39% of values > 0 respectively for variants [a], [b], [c] and [d].

Figure 3. Comparison of the distribution of positive values from each LT-TTO variant reported as the proportion of positive values§



§ Shown as a *proportion* of positive values, rather than as count data, as the study design generated more data using [c] than [a] and [b] – see Table 3.

We investigated this issue of 'acuity' further by examining the number of iterative steps taken to achieve indifference in each of the variants – see Table 9.

Table 9. Average number of iterative 'steps' taken to achieve indifference for health states that are worse, the same and better than dead, by variant

	LT-TTO values		
Variant	**Negative values**	**Exactly 0**	**Positive values**
[a]	6.4	3.3	6.0
[b]	7.0	3.8	4.9
[c]	6.6	4.7	5.8
[d]	7.2	2.0	5.7

Compared to [a] and [c], variant [b] involved a somewhat higher average number of steps to elicit negative values combined with a smaller number of steps for positive values. This suggests that where the length of the available scale for negative values is longer, more steps are taken, and when the length of the scale for positive values 0 is shorter, this lower 'acuity' results in fewer steps. We tested this formally using a regression analysis to examine the effect of scale length for trading (in millimetres), and respondent characteristics, on the number of iterative steps. Scale length exerts a highly significant but small effect: a 1cm increase in scale length increased steps by 0.1. Sex and employment status were both significant - females and those who are employed take an additional 0.6 and 0.8 steps respectively – while age was not significant.

The mean Spearman's rho for each variant indicates a high level of agreement between the ranking of health states implicit in participants' LT-TTO values and their direct ranking of those states, as seen in Table 10. There was little difference between the variants in this respect.

Table 10. Consistency between participants' direct ranking of states and the ranking implied from LT-TTO tasks, by variant

Variant	Spearman's rho (mean)
[a]	-0.79
[b]	-0.71
[c]	-0.75
[d]	-0.74

We examined the prevalence of 'non-trading' i.e. where participants decline to trade off any time in Option A; in effect meaning that the health state under consideration is 'as good as full health'. Rates of non-trading were low, even for the mildest states, across the variants - see Table 11. This was particularly so for the matched variants [c] and [d]. In the lag time TTO (variant d), there was no non-trading.

Table 11. Proportions of non-traders by state and by variant

Health state	Variant a	Variant b	Variant c	Variant d	All variants
11112	3%	3%	1%	0%	**2%**
11211	2%	2%	1%	0%	**1%**
12111	0%	6%	1%	0%	**2%**
11122	0%	0%	0%	0%	**0%**
22121	0%	0%	0%	0%	**0%**
23232	0%	0%	0%	0%	**0%**
33333	0%	0%	0%	0%	**0%**
All states	**1%**	**1%**	**0%**	**0%**	**1%**

Notwithstanding the longer lead times relative to duration in all the variants tested in this study (compared to our earlier research), we observe a non-trivial group of participants who exhaust the lead time in the valuation of severe health states. For example, for state 33333 (which provides the highest *a priori* probability of exhausting lead time), around a fifth of participants in each of variants [a], [b] and [c] did not find the lead time sufficient to express their distaste for this state – see Table 12.

Table 12. Proportions of lead (lag) time exhausters for state 33333, by variant

Variant	LT exhausters
a	21%
b	20%
c	18%
d	28%

The lag time variant [d] had the highest proportion of lead time exhausters for this state. However, a probit analysis of lead/lag time exhausting behaviour (Table 13) shows that [b], [c] and [d] were not statistically significantly different from variant [a] in this respect.

Table 13. Results of a probit analysis of exhausting lead time for state 33333.

Variant	Estimate	Std. Err.	z	P>|z|	95% confidence interval	
Constant	-0.799 *	0.142	-5.64	0.000	-1.077	-0.521
b	-0.043	0.201	-0.21	0.833	-0.437	0.352
c	-0.123	0.184	-0.67	0.505	-0.483	0.248
d	0.204	0.226	0.90	0.366	-0.238	0.646

#observations 414
Log likelihood -208.97
Note: * = significant at 1% level

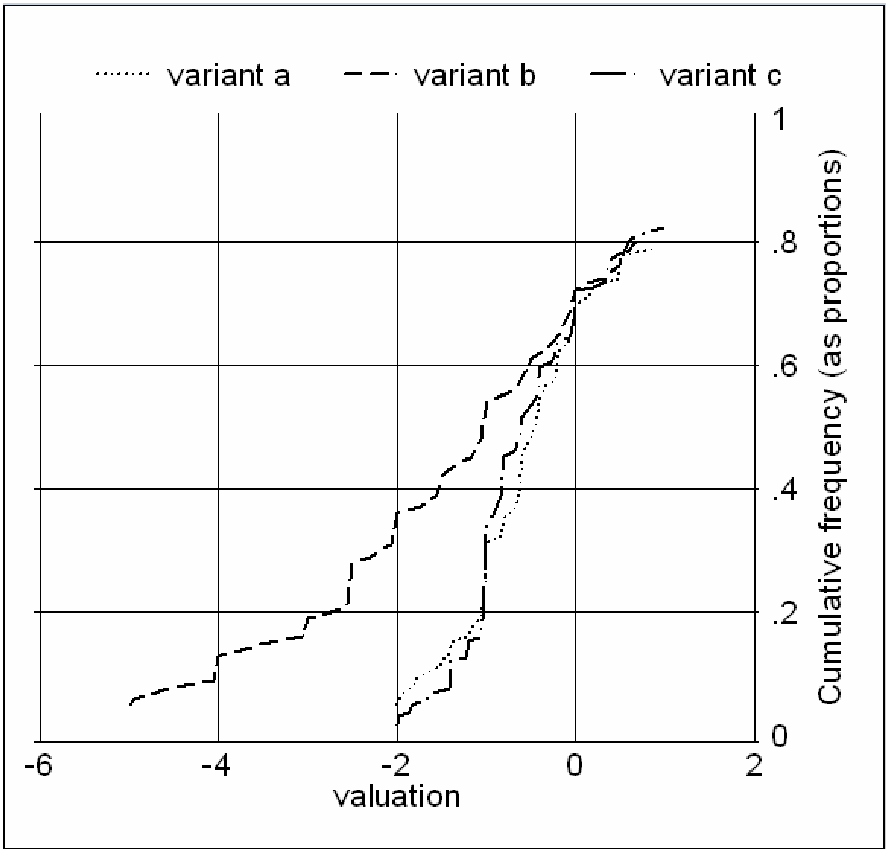
Both lead or lag time methodologies introduce additional time in good health - which effectively increases the *total* utility of the overall prospect (Option B) being valued. The greater the amount of this lead/lag time, the more likely it is that the utility will be positive and hence the greater the likelihood that we can capture preferences (as implied by the equilibrating value of time in good health). We can represent the extent to which preferences are captured, for any given state of health, as a 'preference capture curve'. Table 14 reports the relationship between the length of the lead time, and the proportion of negative values which can be captured *without* resorting to any adjunct procedures (in the case of the conventional TTO, all separate procedures for SWD; in the case of LT-TTO, any additional supplementary tasks used for establishing values when lead time is exhausted).

Table 14. The proportion of preferences for states worse than dead captured by ratio of lead time to duration

Lead time	Duration	Ratio of lead:duration	Preferences captured	Source
0	10	n/a	0%	Dolan (1997)
15 years	10 years	1.5	70%	Devlin et al (2010a)
20 years	10 years	2	76%	Current study
10 years	5 years	2	81%	Current study
5 years	1 year	5	78%	Current study

The distribution of the values for each variant was further examined as cumulative frequencies. For example, Figure 4 shows the cumulative frequency of values for state 33333, without using the data from the supplementary tasks. The distributions of values for variants [a] and [c] are very similar, while the distribution of values for [b] is markedly different from both. While the cumulative frequency of values in all variants are characterised by a 'stepped' rather than a smooth distribution, this was much more marked for [a] and [c], with the drop in the cumulative density function at -1 indicating a large cluster of values centred on that value. Overall, it appears that respondents use a wider range of negative values if that is available. Where the range of values was limited to -2 (variants [a] and [c]) responses were distributed across this narrower range in comparison with variant [b] where a wider range was available. The distribution of positive values 0 is very similar for all lead time variants.

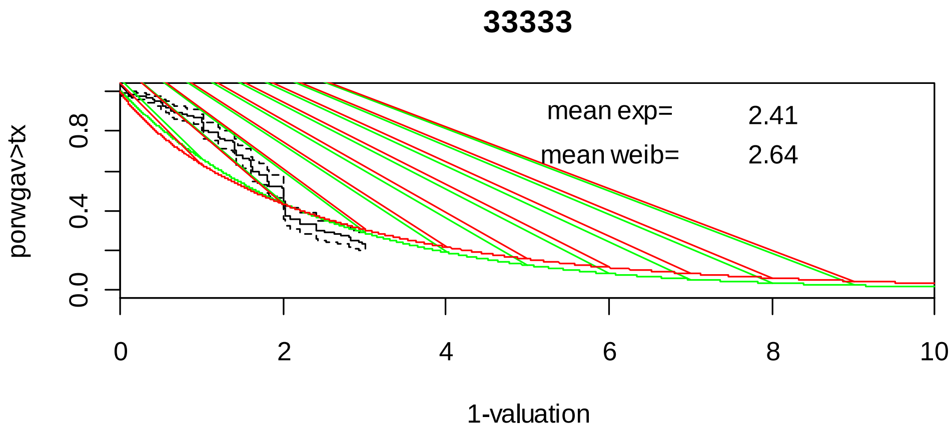
Figure 4. Cumulative frequency of values for state 33333, for variants [a], [b] and [c]

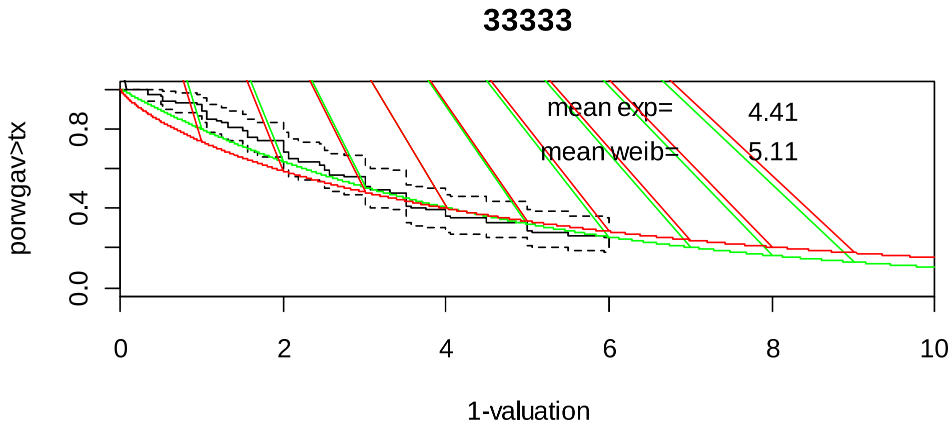


Note: This analysis excludes the data from the supplementary valuation tasks.

Survival analysis was explored as a method to estimate the mean value for health states with censored observations (people who exhausted their lead time). Note that these analyses do not include data from the supplementary tasks. Results are illustrated in Figure 5, showing the cumulative distribution of values for 33333. The upper panel shows the results when using data from variants [a], [c] and [d] (with lead:duration ratio of 2:1) the lower panel when using data from [b] (with a lead:duration ratio of 5:1). It is concluded that the fit is far from perfect and that different survival models may lead to substantial differences in expected values. Most notably the use of these methods illustrates that there may be substantial differences between the results when using different ratios of lead time to duration of the state being valued When using a ratio of 2:1 we find estimates of (1-value) of around 2.5 (implying a value of around -1.5 for 33333). When using a ratio of 5:1 we find value estimates of -3.41 or -4.11 for 33333. More detailed analyses of the fit of the distributions raises questions about whether survival models are to be used in this case. This is further addressed in the discussion.

Figure 5[§]. Parametric accelerated failure time models for EQ-5D health state 33333, with minimum values of -2 (variants [a], [c] and [d]) and -5 (variant [b])





§ The vertical axis measures the estimated share of individuals who have a higher (1-value) than that measured on the horizontal axis. The black lines present the Kaplan Meyer curves (and corresponding 95% confidence intervals); the red line is the estimated Weibull curve; the green line, the estimated exponential curve.

Table 15 reports results from a one-off question that was asked of all participants directly following their valuation of state 33333. The question was 'Do you consider this health state to be better or worse than being dead?' and provided three options: better, worse, or the same.

Table 15. Agreement between participants' declaration of state 33333 being 'worse than dead' and their LT-TTO valuation of it

Declared view of the state	LT-TTO values: Better than dead	Same as dead	Worse than dead	Total
Better than dead	15	0	8	**23**
Same as dead	4	2	34	**40**
Worse than dead	43	4	97	**144**
Total	**68**	**6**	**139**	**207**

This shows that 67% of those who valued this state using the LT-TTO gave it a negative value. Similarly, 69% stated that they thought this state was worse than dead. However, quite frequently these were not the same people. Of those who gave a positive value to this state, just 15 (22%) expressed a view about the state which was consistent with that value, while 43 (63%) contradicted their value by saying that in their view the state was worse than being dead. Of the 139 participant who gave a negative value to this state, just 8 (6%) contradicted that by declaring the state as better than being dead, although 34 (24%) stated that it was the *same* as being dead. Overall, the majority (70%) of those who gave a negative value to a state also expressed the view that the state was worse than being dead.

Table 16 summarises the responses to the structured feedback questions. Participants were asked to specify their level of agreement with each statement using a five-level Likert item (1 = strongly agree; 2 = agree; 3 = neither agree nor disagree; 4 = disagree; 5 = strongly disagree).

Table 16. Tabulated responses to the structured feedback questions.

Statement	Mean	Median	Strongly agree or agree
1. I found it **easy to understand** the tasks I was faced with today	1.5	1	91%
2. I found it **easy to tell the difference** between the health states I was asked to think about	1.7	2	85%
3. I found it **difficult to decide** on the exact point at which I thought Options A and B were the same	2.2	2	65%
4. I found it **difficult to imagine** what it would actually be like to live these 'lives'	2.9	3	44%
5. I found it **easier to complete** the tasks if the **poor health** state lasted for a **longer** period of time	2.9	3	39%
6. The number of years of life in the tasks was **too long** for it to be meaningful to me	3.9	4	13%
7. The number of years of life in the tasks was **too short** for it to be meaningful to me	3.8	4	16%

8.	Tasks involving **longer time horizons** were **harder to imagine** than ones involving shorter time horizons	2.7	3	47%
9.	When I thought about living in the very poorly health states, I took into account the possibility that some **treatment or relief would be provided**	2.9	3	48%
10.	When the yellow state in Option B was very poorly, I thought I could choose Option B, but then choose to die at the point where the yellow state starts	3.1	3	40%
11.	It is not as bad to become ill **in the future** as it is to become ill **now**	2.4	3	60%
12.	When you live in a very poorly state for a long time, it can get more and **more difficult to cope with**	1.6	1	82%
13.	When the yellow state was very poorly, one important consideration for me was whether I would be too much of a burden on my family in that state	2.5	2	61%
14.	When you live in a very poorly state for a long time, you can get used to it and learn to live with the health problems	3.0	3	37%
15.	When completing a task, I **tried to be consistent** with how I had answered the previous questions	1.6	1	88%
16.	I think going from good health to poor health is more realistic than going from poor health to good health	1.8	1	81%
17.	The visual aid in this survey was helpful	1.4	1	92%

Nearly all participants indicated that they found the LT-TTO tasks easy to understand and that the digital aid was helpful. Over 80% participants agreed that considering an option comprising time in full health followed by time in poor health (as in the lead time) is more realistic than going from poor health to good health (as in the lag time). A worrying result, which applies to TTO generally, is that just under half the participants indicated that when they valued the poor health states, they took into account the possibility that some treatment or relief would be provided. This suggests a need to strengthen instructions to participants *not* to take this into account, but to focus on the state as presented, and un-ameliorated.

Similarly, just under half reported that they found it challenging to imagine living in EQ-5D health states that were hypothetical to them – an issue for valuation generally, not just LT-TTO.

Where health states are being valued for longer durations, participants' values may also be coloured by their views about the way that duration will affect their experience of the state. For example, 82% of participants agreed with the statement that "When you live in a very poorly state for a long time, it can get more and more difficult to cope with" and 37% agreed with the statement "When you live in a very poorly state for a long time, you can get used to it and learn to live with the health problems"; each view potentially introduces a bias (operating in different directions) into the valuation task. Neither view was statistically related to participant's experience of serious illness in themselves or in others.

A consideration in choosing between the variants is the plausibility of very long time horizons (as in variant [a]) for elderly participants, and the plausibility of very short time horizons (as in variant [b]) for younger participants. However, there was no statistically significant relationship between age > 60 and agreement with the question "Tasks involving longer time horizons were harder to imagine than ones involving shorter time horizons". There was also no statistically significant relationship between age > 60 and agreement with the question "The number of years of life in the tasks was too long to be meaningful to me". The relationship between age < 30 and agreement with the question "The number of years of life in the tasks was too short to be meaningful to me" approached significance (p = 0.063).

We examined the relationship between participants' responses to these questions and the values they provided by LT-TTO variant; however, as each participant completed *two* variants we considered the results to be uninformative. However, the questions may be of use in future single-variant valuation studies as a means of probing the link between values and underlying views and opinions.

6. Discussion and implications for EQ-5D valuation

LT-TTO yields values for states worse than dead without the need for a separate elicitation procedure or any ad hoc ex-post transformation of data – an important step forward. Recall that the MVH method involves varying both the amount of time in good health and the amount of time in poor health - within a fixed total amount of time. Our primary concern with this method is that when the amount of time in poor health is varied, the thing that is being valued also changes. This is problematic if the utility flow attributable to a given state of health differs according to the duration over which it is experienced, and LT-TTO overcomes this. However, there are some remaining issues with LT-TTO, not least of which is choosing *which* variant of the LT-TTO to select. This choice comprises two related elements: the duration of the state to be valued, and the lead (or lag) time in full health to accompany it.

The choice of health state duration raises, as in any TTO method, the issue of constant proportionality and potential violations of that, such as maximum endurable time (Sutherland et al 1982). As noted in the Introduction, conventional TTO valuation of EQ-5D has historically – although with little real justification – employed a health state duration of 10 years. There are obviously a very large number of possible combinations of alternative health state durations and corresponding lead times (and lag times) that might be considered plausible candidates for LT-TTO. Our study could only test a small number of lead (and lag)-to-duration combinations, but nevertheless provides evidence to help inform the choice of variant.

A striking result is that variant [b] yields mean values that are consistently lower than the other variants. This result appears to be driven by a combination of factors. First, with respect to the *mild* states, while all the variants had a large proportion of values clustered just below 1, in the case of [b] this mode was somewhat lower than the mode of other variants (eg. for state 11112, 0.96, compared to 0.98). While this difference may seem very small, the large number of responses involved drives down the mean. The particular value of the mode in each variant is a direct product of the iterative process used rather than the higher ratio for variant [b] *per se* – and emphasizes the considerable importance of these processes as an influence on values. It is also the case that some respondents in variant [b] considered state 11112 to be far worse than being dead. The fact that variant [b] provided a far wider range of such values in the digital aid scale appears to encourage some people to attribute very low values to the state. This had a large effect in reducing mean values.

The consistently lower means for *severe* health states in [b] is driven by a separate set of factors that seem to be quite complex in nature. It appears that the *ratio* of lead to duration exerts a key influence on negative values: the higher the ratio (5:1 in [b], compared to 2:1 in [a] and [c]) the lower the minimum value (-5 compared to -2). The higher relative amount of lead time available to participants does *not* make them more likely to give a negative values - but where participants *do* value a state as being worse than dead, the greater amount of lead time seems to exert a framing effect that leads to 'more negative' values. In the absence of framing effects, the cumulative frequency curves for variants [a], [b] and [c] in Figure 4 should overlap in the range above -2. Also suggestive is the proportion of those who exhaust lead time for state 33333, which was not much less for variant [b] compared to variants [a] and [c]. However some caution is required: our study does not constitute a 'pure' test of the effect of framing: variant [b] also had the lowest duration (one year) compared to the other variants[6].

One of our study aims was to examine the relationship between the nature of the variants and their ability to capture the disutility associated with the most severe states. While we expected variant [b] to have some advantage in this respect, in practice there appears to be little difference between the proportion of negative values that variant [b] can capture compared with the other variants. Increasing the lead-to-duration ratio to 5:1 increased the proportion of preferences which can be captured, but only marginally.

Other characteristics of the negative values are also worth noting. The distribution of values in both variants [a] and [c] has a marked discontinuity around -1. While there is also a 'step' in the distribution of values at -1 for [b] it is much less prominent. The explanation for the clustering of [a] and [c]'s values at -1 is not obvious. We considered the possibility that this might be attributable to a difference in the average number of 'steps' involved in the iterative procedure before indifference was achieved: there was a somewhat smaller number of steps for negative values in both [a] and [c] compared to [b], and a higher number of steps for positive values; but the difference is quite small.

Where the lead time available within a given variant was exhausted by a participant – meaning their value was lower than the minimum possible in that variant – we experimented with a range of alternative ways of handling that, both by supplementary data collection and by modeling. The former involved follow-up tasks to establish a point of indifference by (i) extending the lead time (keeping duration constant), and (ii) by keeping the lead time the same, but reducing the duration. For state 33333, the lead (lag) time variants captured on average 63% of participants' values; an additional 14% of participants were able to provide values using these supplementary valuation tasks. Relatively few options were offered to participants in these supplementary tasks; the approach could be improved by allowing trades in finer units of time, allowing greater precision in the identification of these values. Both extending lead and reducing duration appeared to be equally feasible as means of identifying values below the minimum possible in LT-TTO variants. However, on conceptual grounds we would argue that extending the lead time is the superior approach. Reducing the duration suffers the same

[6] The idea that it is the ratio rather than the duration that is important is given (weak) support by the similarity in the mean values for variants [a] and [c], which used different durations (10 and 5) but had the same lead: duration ratio. Furthermore, the shorter duration of [b] implies that if maximal endurable time is at play, the values for the severe states should be higher under [b] than under [a] and [c] with longer durations.

problem we have noted of the conventional TTO for states worse than dead: it is varying the 'valuand' (the thing which is being valued).

Regardless of which variant is chosen, we observe that after employing the supplementary tasks, there still remains a small group of participants (about 5%) who cannot achieve a point of indifference in valuing the most severe states. The interpretation of these responses, and how they should be reflected in social value set studies, requires careful thought. One explanation is that they are responding to the health states in a categorically different way – their response to the TTO task may be signaling a qualitative judgment that this is "a very poor health state indeed" (Devlin et al 2010). It is possible that there is *no* TTO approach which will enable these preferences to be quantified. An alternative explanation is that this 5% of participants accidentally box themselves into a 'bravado' response of saying they would not tolerate *any* time in poor health – and these responses to the iterative trade off tasks may or may not tell us anything useful about their true preferences regarding these states.

An alternative way of handling the remaining challenges with these 'extreme negative values' is to focus analysis on median values – where these 'outlier' extreme values become less important – rather than on means. Which measure of central tendency is an appropriate representation of social preferences is as much a normative as an empirical question.

Comparisons between the lead and lag time TTO suggest some complex differences in these approaches. While variants [a], [c] and [d] generated similar mean values for the mild-moderate state, variant [d], the lag-time variant, gave considerably higher values for the severe states compared with its lead-time counterpart and all other variants. This is likely to be related to the observation that variant [d] also had a considerably higher proportion of positive values for those severe states eg. for state 23232, 50% of the variant [d] values were positive, compared to just over 30% in each of variants [a], [b] and [c]. For state 33333, the proportion of positive values in [d], 28% was double that in variant [c]. Given that [c] and [d] are otherwise identical, this difference is purely attributable to the positioning, in Option B, of poor health first and full health later. Maybe the prospect of enjoying a period of full health in the future (in Option B), once the period of poor health is out of the way is preferable to enjoying good health now, with the prospect of poor health 'looming' in the future. However, that argument, and the finding itself, contradict the direction of influence which time preference might have been expected to have. Variant [d] also appeared to completely eliminate 'non-trading' responses (where poor health is valued at 1). That is, confronted with even very mild health states, all 50 participants who were given the lag time task were willing to trade at least *some* time.

Overall, the results suggest the lag time prompts more people to trade (for mild states), but to trade off less time (for severe states). The 'intuition' behind this result may be as follows. When the health state in question is mild, the introduction of lead time shifts the state into the future, and time preference will blur the distinction between full health and the mild state, thus leading to non-trading. However, with the introduction of lag time, the health state in question starts now, so the distinction between it and full health is not blurred by time preference. In addition, because the state is mild, violation of additive separability is not a major issue. In contrast, when the health state in question is severe, violation of additive separability may take over. To the extent that people prefer a health profile that is improving through time rather than worsening, a severe state with lag time will be more attractive than a severe state with lead time. Furthermore, if the health state is distinctly different from full health, then time preference will not cause non-trading.

A key question for the LT-TTO is the extent to which participants who provide negative values are also of the opinion that the state is worse than being dead. Our study tested this directly. Most of those (70%) who valued state 33333 as negative also expressed the view that the state was worse than dead. A further 24% stated a view that the state was about the same as being dead, while just 6% exhibited a rather stronger sort of inconsistency by expressing the view that the state was in fact *better* than dead. Similarly, of those who expressed the view that health state 33333 was worse than dead, 67% provided LT-TTO values consistent with that. While the results are far from unequivocal, they do provide some support for the legitimacy of negative values in LT-TTO. There are no directly comparable data to these available from other TTO studies.

The use of a digital aid conferred important advantages in this study. Compared to physical props (TTO boards) digital aids could readily operationalise the different variants of the tasks. Participants reported that they found the digital aid helpful. The use of a digital aid appears to speed up the interview process considerably: the average time taken per interview in the pilot (comprising 9 TTO tasks plus other tasks) was 36 minutes and in the main study 24 minutes (comprising self-reported health, ranking of 12 states, 10 TTO tasks plus background and feedback questions). This compares to an average of 54 minutes taken per interview (sd 15) in the MVH study (comprising self-reported health, ranking of 15 EQ-5D states, TTO of 13 EQ-5D states and background questions)

The shorter interview times in the main study compared with our pilot might arise for a number of reasons (eg. the former was conducted by professional interviewers who gain experience by conducting many interviews within this study, rather than by the research team who each conduct a relatively small number of interviews). However, a key difference between the pilot and main study was that in the latter, the iterative process and data capture were fully automated. We think it is likely that this is a key factor in the speed of the interviews, as the interviewers do not have to pause to consider what iteration should follow a given response, and do not need to record those responses. This may suggest it is possible to increase the number of LT-TTO tasks which can be done by each participant, with implications for sample sizes and costs. However, a potential concern is that digital aids may make it possible to complete the valuation tasks *too* quickly: participants may be afforded less opportunity to reflect on the state in question, and their responses to the trade-offs presented. However, notwithstanding some short interview times, these did not appear to be associated with any observable or systematic deterioration in the quality of the data. In addition to speeding up data collection, the digital aid also facilitated the capture of information on the iteration process, particularly the number of steps taken to achieve equilibrium.

When respondents exhaust their lead time, they implicitly state that the value of that health states is less than the minimum possible within the LT-TTO task. As such the information is censored and survival models were used as a first way of handling this. As noted earlier, to our knowledge this is the first study to use survival analysis in the context of stated preferences research. As such, distributions which are normally used to describe the distribution of survival times are now used to describe the distribution of 'opinions' (health state preferences). The imperfect fit of these models suggest that there are some challenges in applying survival analysis to preferences data; and more complex survival curves do not appear to solve this problem. Unfortunately, whereas increasing and decreasing hazard rates make perfect sense in survival times, they do not when describing distributions of opinions. Other, more tailored methods, also using maximum likelihood functions (capturing the censored information) could be explored. However, all this may not solve the problem we note above: that there may always be some individuals who state that some health states are unacceptable at any price and in any trade off, suggesting values approaching minus infinity.

Finally, there remain a number of research questions which our study has not addressed. First, time preference potentially exerts an effect on valuations, an issue that continues to be of relevance to all TTO-based valuation methods, and particularly so to LT-TTO variants involving longer time profiles. Second, the development and testing of LT-TTO valuation methods that have been undertaken so far and reported here all focus on EQ-5D (i.e. 3-level) states. Ultimately, however, the valuation methods now being developed will be applied to the valuation of EQ-5D-5L states. The additional number of levels within dimensions, and the greater subtlety of the labels, may pose additional challenges for participants in differentiating between and valuing states in LT-TTO. The valuation of EQ-5D-5L health states is currently being investigated in research underway in England and the Netherlands.

Acknowledgements

We are grateful to all interviewees who took part in our study. Ethics approval for this research was obtained from the University of Sheffield. This study was funded by Department of Health Policy Research Programme grant PRP 070-0065. Support for the pilot study was provided by the EuroQol Foundation. Views expressed in the paper are those of the authors, and not the Department of Health. We are grateful to Professor Martin Buxton for comments on an earlier draft.

References

Buckingham K, Devlin, N. (2009) An exploration of the marginal utility of time in health. *Social Science and Medicine* 68: 362-367.

Chevalier J, de Pouvourville G. (2009) *Valuing EQ-5D using TTO in France*. ISPOR 12th Annual European Congress.

Craig BM, Oppe M.(2010) From a different angle: a novel approach to health valuation. *Social Science and Medicine* 70(2):169-174

Devlin N, Tsuchiya A, Buckingham K, Tilling C. (2010) A uniform Time Trade Off method for states better and worse than dead: feasibility study of the 'lead time' approach. *Health Economics* (forthcoming).

Dolan P. (1997) Modelling valuations for EuroQol health states. *Medical Care* 35(11):1095-108.

Kind P, Hardman G, Macran S (1999), *UK Population Norms for EQ-5D*, Centre for Health Economics, University of York

Lamers L, McDonnell J, Stalmeier P, Krabbe P, Busschbach J. (2006) The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Economics* 15(10):1121-32.

MVH Group. (1994) *The Measurement and Valuation of Health: First Report of the Main Survey*, Centre for Health Economics, University of York.

National Institute for Health and Clinical Excellence. (2008) *Guide to the Methods of Technology Appraisals*. Publication reference N1618. London: NICE.

Office for National Statistics. (2009), http://www.statistics.gov.uk/cci/nugget.asp?id=6; accessed 14.06.2010

Robinson A, Spencer A. (2006) Exploring challenges to TTO utilities: valuing states worse than dead. *Health Economics* 15: 393-402

Shaw J, Johnson J, Coons SJ (2005) US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Medical Care* 43(3): 203-20.

Szende A, Oppe M, Devlin N. (eds) (2007) *EQ-5D Value Sets: Inventory, Comparative Review and User Guide.* Springer.

Sutherland HJ, Llewelyn-Thomas H, Boyd NF, Till JE. (1982) Attitudes towards quality of survival: the concept of “maximum endurable time”. *Medical Decision Making* 2:299-309.

Tilling C, Devlin N, Tsuchiya A, Buckingham K. (2010) TTO Valuations of Health States Worse than Dead: a literature review and conceptual framework for systematic analysis. *Medical Decision Making*. 30(5):610-619.

Tsuchiya A, Ikeda S, Ikegami N, Nishimura S, Sakai I, Fukuda T, Hamashima C, Hisahiqe A, Tamura M. (2002) Estimating an EQ-5D population value set: the case of Japan. *Health Economics* 11(4):341-53.

Verschuuren M. (2006) *Quality adjusted life years and time trade off exercises: exploring methodology and validity.* PhD thesis: University of Utrecht, The Netherlands.