View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by City Research Online

Devlin, N., Tsuchiya, A., Buckingham, K. & Tilling, C. (2009). A uniform Time Trade Off method for states better and worse than dead: feasibility study of the ‘lead time’ approach (Report No. 09/08). London, UK: Department of Economics, City University London.

CITY UNIVERSITY LONDON

City Research Online

Original citation: Devlin, N., Tsuchiya, A., Buckingham, K. & Tilling, C. (2009). A uniform Time Trade Off method for states better and worse than dead: feasibility study of the ‘lead time’ approach (Report No. 09/08). London, UK: Department of Economics, City University London.

Permanent City Research Online URL: http://openaccess.city.ac.uk/1492/

Copyright & reuse

City University London has developed City Research Online so that its users may access the research outputs of City University London's staff. Copyright © and Moral Rights for this paper are retained by the individual author(s) and/ or other copyright holders. Users may download and/ or print one copy of any article(s) in City Research Online to facilitate their private study or for non-commercial research. Users may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. All material in City Research Online is checked for eligibility for copyright before being made available in the live archive. URLs from City Research Online may be freely distributed and linked to from other web pages.

Versions of research

The version in City Research Online may differ from the final published version. Users are advised to check the Permanent City Research Online URL above for the status of the paper.

Enquiries

If you have any enquiries about any aspect of City Research Online, or if you wish to make contact with the author(s) of this paper, please email the team at publications@city.ac.uk.

School of Social Sciences
CITY UNIVERSITY LONDON

Department of Economics

A uniform Time Trade Off method for states better and worse than dead: feasibility study of the ‘lead time’ approach

Nancy Devlin[1]
Office of Health Economics

Aki Tsuchiya
University of Sheffield

Ken Buckingham
University of Otago

Carl Tilling
University of Sheffield

Department of Economics
Discussion Paper Series
No. 09/08

---

[1] Office of Health Economics,12 Whitehall, London SW1A 2DY,Telephone: ++44 (0)20 7747 8858, Fax: ++44 (0)20 7747 8851, Email: *ndevlin@ohe.org*

A uniform Time Trade Off method for states better and worse than dead: feasibility study of the ‘lead time’ approach

Nancy J. Devlin[1], Aki Tsuchiya[2,3], Ken Buckingham[4], Carl Tilling[3]

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

Correspondence to:

Professor Nancy J. Devlin
Office of Health Economics
12 Whitehall
London SW1A 2DY
Telephone: ++44 (0)20 7747 8858
Fax: ++44 (0)20 7747 8851
Mobile: 07 515 974 978

Short running title: Feasibility study of the ‘lead time’ TTO
Key words: TTO; constant proportionality; states worse than dead
Word count: 5,534
Number of Tables: 4
Number of Figures: 3
Funding: This study was supported by a grant from the EuroQol Foundation.

A uniform Time Trade Off method for states better and worse than dead: feasibility study of the 'lead time' approach

Summary

The way Time Trade Off (TTO) values are elicited for states of health considered 'worse than being dead' has important implications for the mean values used in economic evaluation. Conventional approaches to TTO, as used in the UK's 'MVH' value set, are problematic because they require fundamentally different tradeoffs tasks for the valuation of states better and worse than dead. This study aims to refine and test the feasibility of a new approach described by Robinson and Spencer (2006), and to explore the characteristics of the valuation data it generates. The approach introduces a 'lead time' into the TTO, producing a uniform procedure for generating values either >0 or <0. We used this lead time TTO to value 10 moderate to severe EQ-5D states using a sample of the general public (n=109). We conclude that the approach is feasible for use in valuation studies, and appears to overcome the discontinuity in values around 0 evident in conventional methods. However, further research is required to resolve the issue of how to handle participants who 'use up' all lead time; to develop ways of controlling for individual time preferences; and to better understand the implications for valuations of states better than dead.

Word count: 200

Uniform method of Time Trade Off for states better and worse than dead: feasibility study of the ‘lead time’ approach

1. Introduction

The Time Trade Off (TTO) (Torrance et al, 1972) is a widely accepted approach to the valuation of health related quality of life in the estimation of Quality Adjusted Life Years (QALYs). While published papers on TTO are dominated by studies using it to seek patients’ valuations of their own quality of life (Tilling et al, 2008), arguably the most significant use of TTO is in generating the values for ‘generic’ health states commonly used in economic evaluation. For example, the TTO has been used to generate the UK, US, Japanese and Dutch value sets for the EQ-5D, a generic health state instrument (Szende et al, 2007). The UK’s TTO value set for the EQ-5D is recommended by the National Institute of Health and Clinical Excellence for use in evidence submitted to it (NICE, 2008).

The basic TTO approach involves finding the number of years (x) in full health that is equivalent to t years in H_i, at which point $U(H_i) = (x/t)$. The particular protocol for eliciting TTO valuations of EQ-5D health states used in the UK Measurement and Valuation of Health (MVH) study (Dolan, 1997) is widely used in part or full by other researchers. For example, it was used in the Netherlands (Lamers et al, 2006) and the US (Shaw et al, 2006). In the MVH study, participants were asked to imagine that all the health states to be valued would be experienced for t =10 years. If the health state is very severe, participants will be prepared to trade off a large amount of time in full health, yielding a low value. And in the case of very extreme states of pain and other problems, participants may consider experiencing these for 10 years so distasteful that they would prefer immediate death. This creates a challenge: how can TTO values for these states, which the interview process has revealed to be ‘worse than being dead’, be elicited?

The MVH protocol requires participants whose responses indicate States Worse than Dead (SWD) to engage in a different sort of TTO task. This involves choosing between immediate death, and spending a length of time (10-x) in H_i (the state to be valued), followed by x years in full health. The value of x is varied until the participant is indifferent between the two options. This procedure is illustrated in Figure 1. The worse a given health state H_i, the more time is required in full health to compensate for the time spent in H_i. The value for H_i is given by $U(H_i) = -x/(10-x)$. For example, if the participant is indifferent between the option of immediate death, and the option of 6 years in H_i followed by 4 years in full health, then the value of $U(H_i) = -4/(10-4) = -4/6 = -0.67$.

There are important problems with this approach to valuing SWD. The trade-off procedure employed for SWD is fundamentally different, both conceptually and operationally, from that used for states *better* than dead (SBD). Conceptually, the TTO values for SBD, (x/t) are obtained by varying x while t is fixed. In contrast, the TTO procedure for SWD involves *simultaneously* changing both the numerator and the denominator – this has the effect of exaggerating the effect on utilities of the trade-offs made.

Operationally, the use of different elicitation procedures for SBD and SWD means that at the point during the elicitation procedure where it becomes evident the valuation for a given state is negative, the interviewer has to switch to the use of a completely different 'prop' and the participant has to re-engage with an entirely different sort of task. Given these differences, and other problems discussed below, the aggregation of positive and negative values for any given state in order to calculate mean values or to estimate a value set is of questionable validity.

Further, the MVH elicitation procedure produces 'extreme' negative values; and how negative these valuations may become depends on the (relatively arbitrary, and seemingly innocuous) researcher decision about the 'units' of time that can be traded. In the MVH study participants could make tradeoffs in units of 3 months (0.25 years); hence the worst possible health state had a value of -9.75/ (10-9.75) = -39. If the trade-off was in whole years then the minimum value would be -9/(10-9) = -9; or, if it was in months then the minimum value would be -9.16/(10-9.16) = -119, and so on.

Because the elicitation procedure produces such extreme negative values, researchers have responded by doing *ex post* transformations to bound negative valuations to -1 in various ways (Lamers, 1997). Crucially, once transformed, the negative numbers for SWD can no longer be interpreted as “utility” scores, measured on the same scale as those for SBD (Patrick et al, 1994). Yet standard practice in calculating QALYs is to treat all values reported in value sets as commensurable. For example, an improvement from -0.2 (a SWD) to 0, experienced over one year is interpreted as, producing a gain of 0.2 QALYs, and this is treated in technology appraisals as identical to an improvement from 0 to 0.2 experienced for one year - whereas the underlying ‘untransformed value’ for the SWD might suggest these two improvements in health are valued quite differently. This has serious implications for organizations such as NICE that routinely use the MVH TTO value sets in technology appraisals and decision making.

It has also been observed that there is a ‘gap effect’ in the MVH TTO values around dead. In an analysis of differences in the values for adjacent states, Stalmeier et al (2005) find the differences in TTO values for states either just above or below 0 are at least twice as large compared to the differences between other adjacent states. Further, there are few states in the MVH TTO value that are close to 0 (either SBD or SWD). This discontinuity means, for example, SWD tend to be *substantially* worse than dead. This ‘gap’ is probably related to the separate valuation procedure used for SWD. For example, it could be caused by the explicit questions asked of participants regarding whether the state in question is indeed better or worse than dead. The notion of a SWD may be disturbing to members of the public, and having declared a state as being so, there may be a focusing effect which amplifies how *much* worse the state is than dead.

The problem of TTO valuations of SWD is not unique to the EQ-5D. Other research groups that use the TTO face the same issues, although the response in each case – both the specific elicitation procedures employed, and *post-hoc* manipulations of the data – differ (for example, see Richardson and Hawthorne, 2001). Neither is the problem restricted to TTO: the Standard Gamble elicitation procedure must also be changed for SWD (see Drummond et al, 2005). The problem of extreme negative values affects both SG and VAS, and should researchers choose to transform the negative values to be bounded by -1 then, similarly, the resulting values will be incommensurable with values for SBD.

A related issue is whether or not values of negative states *should* be bounded to -1. It is not obvious why there should be *no* states worse than -1. For example, the phrase ‘it would have been better if he had never been born’ could truly be applied to people who have undergone torture and other types of brief but extreme suffering. There is no theoretical basis for imposing a limit on the level of disutility associated with these extreme sufferings.

If, as is the case, we interpret SBD as states that contribute positively to one’s stock of lifetime QALYs, then we may correspondingly interpret SWD as those that contribute *negatively* to one’s stock of lifetime QALYs. If this is the case then, for example, all the following scenarios are equivalent in QALY terms: Live to 50 in full health and die; *or* Live to 51 in full health, then live in -1 for 1 year, and die at 52; *or* Live to 52 in full health, then live in -2 for 1 year, and die at 53; *or* Live to 53 in full health, then live in -3 for 1 year, and die at 54. This suggests that values less than -1 may be both conceptually plausible and represent meaningful expressions of value. If this point is conceded, then the justification for transformation is greatly reduced - while increasing the importance of designing a method of eliciting negative valuations that are meaningful.

This study aims to develop and test the feasibility of a new TTO approach – the ‘lead time’ TTO (described below) – that overcomes the issues with the MVH, and to explore the characteristics of the EQ-5D valuation data it generates. In doing so, maximum comparability with the MVH design was maintained.

2. The lead-time TTO

We undertook a critical review of all studies which had elicited TTO values, to identify how SWD had been handled and to evaluate the merits of alternative procedures – see Tilling et al (2008). A number of alternative approaches to TTO were considered and rejected. For example, Torrance's original TTO procedure for SWD (Torrance, 1982) differs from the MVH only in the order in which full health and H_i appear in the task - for example, in Figure 1, Torrance's approach would present Life A as comprising full health then H_i, rather than H_i to full health as in the MVH. The change was introduced in the MVH study in order to make sure participants' values were not biased by their considering the possibility of committing suicide at the end of the period in full health In the present study, the Torrance approach was rejected on the grounds that it suffers the same problems as the MVH protocol.

Robinson and Spencer (2006) report an approach comprising a series of choices using a set of visual aids, which in effect can be interpreted as a TTO procedure that can be applied uniformly for both SWD and SBD. Participants ranked a series of life profiles consisting of combinations of good health, poor health and death. The principal innovation of the method is its introduction of a 'lead time' in full health preceding each of the alternatives presented – see Figure 2. The approach avoids the need to have different valuation procedures for SBD and SWD by allowing participants to trade their lead time to avoid states considered worse than dead. We considered this the most promising of the alternatives we had identified, and selected it as the basis for our study.

In order to calculate values from the lead time TTO responses, the lead time is subtracted from both the numerator and the denominator to give a result comparable with the usual TTO. Where the lead time is 10 years, $U(H_i) = (x - 10)/20-10$). If at the point of indifference, 'x' years in full health in Life A is *greater* than the lead time in Life B (see Figure 2), the value will be positive; if it is equal the value is equal to 0, and if x is *less* than the lead time in Life B, the value will be negative. Thus the derivation is identical for SBD and SWD.

3. Protocol design

An interview protocol and physical prop was developed to implement the lead time concept, but mimicking the MVH design i.e. using a TTO board fitted with a sliding device, and using similar colours for the health states, to illustrate the trade-off and to allow iteration toward a point of indifference. The interview protocol and TTO board were developed through a process of preliminary testing and refinement, and then piloted on a convenience sample of 27 administrative staff at City and Sheffield Universities, using just three EQ-5D states.

The pilot incorporated experimentation with alternative durations both for the health states and the lead times. The TTO board used in the pilot was designed in such a way that the front-piece could be removed and replaced with another to show different lengths of time. The basic design involved a lead time of equal length to the duration of the state being valued (for example, 10 years lead time followed by 10 years in H_i). Where this lead time was ‘used up’, we switched the front-piece and continued the TTO using a longer lead time. Further details of the pilot study are reported in Devlin et al (2009).

The pilot study results suggested that the alternative duration TTOs were equally feasible; a 10 year duration was selected for the feasibility study reported here on the grounds that this would facilitate comparisons with the MVH value set elicited for the same duration. The interview protocol and script used in the pilot was further refined e.g., by making the ‘branching’ instructions clearer.

Each participant in the feasibility study was asked to value 10 EQ-5D states. For comparison, the number of states per participant included in EQ-5D TTO valuation studies was 9 in Zimbabwe; 13 in the UK, US, Spanish and German TTO studies; 16 in the Denmark study; 16 and 17 in the Japanese and Dutch studies (Szende et al, 2007). The 10 states were selected from the set of states evaluated in the MVH study; but drawing principally on severe states where we were more likely to observe responses indicating SWD.

Prior to the TTO exercise participants were asked to rank the 12 states (the 10 intermediate states, plus full health, and dead). Because ranking 12 states was regarded as somewhat time-consuming and demanding, they were presented in pre-determined batches of four. Once they had ranked the first four, participants were handed the second batch of four cards and asked to place them where they thought they belonged in the ranking; likewise for the following sets. After the ranking, participants placed the states on a standard Visual Analogue Scale (VAS). The position of the cards on the scale was determined by the ranking exercise: the first ranked state was placed at the top of the scale and the 12th ranked state was placed at the bottom of the scale. This was to minimise the crossing of lines when the participant drew from the box to the scale, which subjects sometimes show a reluctance to do.

In the pilot study we found a number of participants ‘exhausted’ (used up) all their 10 year lead time when contemplating the most severe state, requiring the use of the extended (20 year) lead time board front-piece. For the feasibility study, it was deemed preferable to avoid the requirement to do this by increasing the lead time to 15 years.

Given the possibility that even the 15 year lead time might be exhausted, the protocol was designed so that if this arose, participants were asked a simple binary yes/no question about whether they would still prefer Life A (in effect, as the lead time has been traded away at that point, immediate death) if Life B now comprised 20 years in full health and 5 years in *Hi*. Note that this revised Life B maintains the same overall duration (25 years) but involves a reduced duration in *Hi* and thus a change in the denominator for calculating the health state value (from 25-15=10 to 25-20=5). An alternative would have been to revise Life B to be a longer lead time (e.g., 25 years) in full health followed by 10 years in *Hi*, which would maintain the same denominator but involve an increase in the total duration of Life B, and thus a delay in the age of death. Neither approach is ideal; we return to this point in the Discussion.

After the TTO exercise, there was a series of structured-response feedback questions to capture participants’ views about the valuation tasks. These were developed based on the responses to open ended feedback questions in the pilot study. We also attempted to gauge each participant’s rate of time preference by asking a short series of questions that involved weighing up the duration and timing of headache episodes. The interview ended with questions on respondent background characteristics. In addition to the more standard questions used in similar surveys, we included a question on after-tax household income. The categories used were based on income quintiles from the Office of National Statistics for 2006-2007. The interview protocol and further details of the TTO board are available from the authors on request.

4. Data collection and methods

Interviews with members of the general public were conducted by a team of four interviewers at Sheffield Hallam University (SHU). The interviewers had previous experience in eliciting TTO values for hypothetical health states, and were given training and practice in using the lead time TTO prop and interview script. The sample was selected using SHU's standard recruitment frame, "AFD software names and numbers". A random sample of areas was chosen, based on postcode, in the South Yorkshire area. A letter inviting participation and an information sheet on the project was mailed out to addresses in those areas in February 2008. Interviewers then called on the houses in those areas in the following four weeks, to achieve a target of 100 interviews.

The background characteristics of the sample and the rank, VAS and TTO results are all compared against the MVH study results. For the rank results we determine the average rank for each, assuming rank scores are cardinal. We calculate the percentage of respondents giving a particular rank to a given state. This distribution is then used to determine the modal rank. Finally, rank ordered logit regressions are used to predict the probability of a given state being ranked first.

Participants' VAS valuation data are excluded from the analysis if they are implausible (for example, dead valued > or = 11111). The VAS results are re-scaled so that 0 represents dead and 100 represents full health (11111). This is done using the transformation (H_i - Dead)/(11111 - Dead). We present the mean, median and standard deviations for our raw VAS results, the MVH raw VAS results, our re-scaled values and the MVH re-scaled values.

TTO values were calculated as $U(Hi) = (x\text{-}15)/(25\text{-}15)$. Regarding the binary yes/no question asked where lead time was exhausted, if participants preferred Life B (20 years' lead time plus 5 years in H_i), their value was set at -1.5. If they still preferred Life A (immediate death) these were coded as (0-20)/(25-20) = -4 and these are reported as the 'baseline' values for those responses. However, the reliability of this last response may be questioned, as it was not based on the purpose built visual aid, neither was an equilibrating value sought. Therefore, to check the sensitivity of our findings to this treatment, we also report the results from an alternative treatment of the responses where the 15 year lead time was exhausted, by giving them an artificial value of -1.5. Furthermore, in order to facilitate comparisons with the MVH values (which are transformed to -1), a third set of TTO values were generated by dividing all negative values under the baseline approach by 4, so that the minimum value becomes -1.

If the only impact of the introduction of lead time is the way in which SWD are valued, then it should not affect whether or not a given state is valued better or worse than dead. Therefore, the proportion of respondents regarding each state as worse than dead is compared with the MVH data using Chi^2 tests. Furthermore, while the lead time TTO is expected to affect the distribution of negative observations, it should not affect the distribution of positive observations. This is explored by comparing the mean and the quartiles of positive observations of the feasibility study with the MVH data, and by conducting two-sample Kolmogorov-Smirnov tests by state. This is a non-parametric test of equality of distributions.

Regression models are used to study the effect of background characteristics on the rank, VAS and TTO results. To look at the effect of background characteristics on the rank results a rank ordered logit model is used. In the case of the re-scaled VAS and baseline TTO results standard OLS regression models are used to test the effect of background characteristics. Coefficients and t-statistics from these regressions are presented. Chi-squared tests are used to investigate the correlation between background characteristics and responses to the feedback questions.

Consistency was examined using an approach in line with the MVH approach: two-way correlations were derived, using each individual's responses, between the following variables: health state rankings, TTO score and VAS score. These were correlated using Spearman's rank correlation method.

To test for interviewer bias we estimate an equation to model the impact of health state on TTO valuations and include dummy variables to represent the interviewers.

Data were analysed using STATA Version 8.

5. Results

Participant characteristics

Interviews were completed by 109 members of the general public. Table 1 shows the background characteristics of the sample, with the corresponding MVH study results for comparative purposes. Our sample, like the MVH sample, contains more females than males. Compared with the MVH sample, our sample contains proportionally fewer people over the age of 64, more people between the ages of 45 and 65, more employed people and fewer retired people.

A smaller proportion of our sample had experience of serious illness both in themselves and in their family, but a higher proportion had experienced serious illness in caring for others. Mean self-reported health on the EQ-VAS is slightly lower amongst our sample, despite the fact that a larger proportion of our sample is employed/self-employed. The income category £20,001 - £26,000 is under-represented in our sample, while the category £26,000 -£40,000 is over-represented. Our sample has a very high level of home ownership. The areas in which the respondents were recruited from were coincidentally all dominated by mortgaged properties rather than rentals.

Ranking, VAS, and TTO results

One of the interviewers initially misunderstood the ranking procedure, where EQ-5D states were presented in batches. Instead of ranking the second and third batches of states alongside the first batch of ranked states, this interviewer conducted three separate ranking exercises. This affected 17 respondents; these participants were dropped from the ranking analysis, leaving 91 participants. State 11111 has the highest and Dead the lowest average rank: no states are, on average, ranked as worse than Dead, including the worst possible state, 33333. The rank order correlation between average rank, modal rank, and predicted probability of a given state being ranked first based on the rank order logit regression, are very high, suggesting that the results of the ranking exercise are robust.

Four respondents were excluded from the analysis of VAS values. One of these gave Dead a value of 100, one gave Dead a value of 98, one gave state 11111 a value of 0 and another gave three states identical values (one of which was Dead). This left a total of 104 responses. Table 2 shows both the raw and re-scaled VAS values for both our study and the MVH study. The VAS values in our study are consistently higher than those in the MVH study. Furthermore, none of our mean re-scaled values are negative, while two states receive a negative value in the MVH study (33323 and 33333).

A total of 109 participants provided the valuation data reported in Table 3. No TTO data were excluded from the analysis; four participants did not value all the states but their valuation data was included.

Caution is required in interpreting these results. Our study was not designed to elicit a valuation set. The lead time is exhausted by at least one respondent in all but one of the 10 states. In the case of the worst possible EQ-5E state (33333), 16 participants (15%) exhausted the lead time. For the more severe states of health, lead time TTO values are lowest for the baseline case, where minima of -4 are used, and become progressively higher when the minima are -1.5 and -1. In the latter case they generally exceed the values derived for the MVH study. Except for the mildest state (11112) the mean MVH values are lower than the lower 95% confidence interval for our results with minima -1.

Table 4 reports additional analysis of the distribution of TTO values > 0 and for those < 0, and how these compare with the MVH. A total of 100 participants provided TTO results for all 10 states, and their results are reported here. The proportion of participants valuing a given state as being worse than dead is statistically significantly different from the MVH study (Chi^2-test, 2-sided, at 5%) in three out of the 10 states: 13332, 11112, and 23232. There is no consistent pattern to this difference – for 13332 and 23232 the lead time approach has produced a lower percentage of values worse than dead, whereas the opposite is the case with state 11112. Table 4 also illustrates how the mean and the quartiles of positive observations compare across the two studies. The Kolmogorov-Smirnov test suggests that the distribution of positive observations are statistically significantly different between the present study and the MVH in four out of the 10 states (the above three states, plus 22222). In each of the four states, the mean value of the TTO values > 0 is higher in the lead time TTO than the MVH. It may be noted that none of these states are at the extreme end of severity in the EQ-5D descriptive system.

Figure 3 shows the distribution of baseline TTO values in our study compared to those from the MVH Study. The gap between -1.5 to -4 in our study is an artefact of the binary yes/no question we asked participants that exhausted their lead time. The peak that occurs at 0 in the MVH study is not present in our lead time TTO results.

The effect of background on ranking, VAS and TTO

A cross-correlation matrix of key variables (results available from the authors) showed that only the correlation coefficient for VAS and ranking exceeded 0.7. Other correlations (coefficients of at least 0.2) were between: TTO and VAS scores; TTO and Rank scores; own ill health and having a degree; own ill health and own VAS score; experience of health in others and experience as a carer; being a home owner and income greater than £26,000; and being older than 45 and being a mother.

None of the participants' characteristics appear to affect the ranking of a given state. However, those who reported having experienced illness themselves or as a carer gave lower VAS values to health states. Similarly, males and homeowners gave statistically significantly lower VAS valuations. With respect to TTO values, having personal experience of ill health, being in work, having a degree, being a home owner, being a female in a household containing children and VAS for one's own current health, are all associated with significantly higher TTO values .

Analysis of participant's responses to structured feedback questions showed that those with children were slightly more concerned about being a burden on their family. Those in worse health found it easier to imagine a given hypothetical state. Those who found it easier to imagine the states showed greater levels of inconsistency between VAS and TTO scores. Participants who considered their ability to work in a given state also exhibited greater inconsistency between VAS and TTO.

Consistency across valuation methods

The mean Spearman correlation coefficient between ranking and TTO was -0.618 (the sign is negative, because low rank value corresponds to high valuation) and between VAS and TTO was 0.582. These values suggest high levels of consistency between the three valuation methods. Note that since the states valued were generally very similar in severity, testing for consistency becomes more difficult.

Time Preference

Results from questions which attempted to elicit time preference failed to achieve useful results: 68 respondents gave inconsistent responses to the four questions which made it impossible to use these data to calculate a rate of time preference for them. Of these inconsistent responders, 47 gave seemingly random responses; 16 stated that all scenarios were equal and five ranked them in reverse. Of the 40 consistent responders, the most prevalent level of time preference was >10% (12 responses).

6. Discussion and conclusions

The issues surrounding TTO valuation of SWD may appear to be a highly specialised topic, of little practical importance. To the contrary, addressing these issues is central to the continued application of TTO values in economic evaluation. Nearly one third of the 243 health states described by the EQ-5D descriptive system have negative values in the MVH value set used by NICE. Further, recall that the *mean* values for *most* health states are a product of values both positive and negative. Moreover, 32 of the 42 states valued in the MVH study had negative mean values prior to the *ex post* artificial transformation to -1. The way in which SWD are handled in TTO valuations is therefore highly relevant to their use in economic evaluation and providing a defensible basis for resource allocation in a publicly funded health care system. In this study, we have identified an alternative approach to TTO which uses a uniform procedure regardless of whether or not a state is worse than dead, and operationalised this procedure maintaining maximum comparability with the established MVH TTO protocol.

Our results suggest that the lead time TTO is feasible for participants. The interviewers reported that the protocol was straightforward to administer. The values elicited using the lead time TTO had a high level of within-respondent consistency with the ranking and VAS evaluations of the states. However, as has been observed in other valuation studies, there was almost near universal positioning of Dead (both in the ranking exercise and in VAS) as the worst state by our sample, in contrast to their subsequent TTO valuations. One explanation is that the ranking of Dead as worst is an initial ‘gut reaction’ to what is the first task in the interview, and that this is modified once the participant engages with the task of imagining themselves to experience these states. However, the same phenomenon regarding the ranking of Dead has been observed when TTO tasks preceded rank/VAS (Tsuchiya et al, 2006). An alternative explanation is that participants might think of rank/VAS as some chronological or logical sequence: we get less healthy, and then we die, so dead is placed at the bottom (Robinson et al, 1997).

There is *prima facie* evidence that the use of a uniform procedure for SBD and SWD overcomes discontinuities in values around 0. By employing a uniform procedure for states better and worse than dead, participants' responses can readily 'flip' from positive to negative values without the focusing effect created by the introduction of a separate valuation procedure or indeed any explicit mention of that state as being worse than dead. It would be interesting, in future research, to ask participants whose TTO responses indicated SWD whether they agreed with that conclusion or indeed were aware that this is what their responses implied.

While the lead time TTO appears to have the potential to overcome the problems of conventional TTO in valuing SWD, its use relies on the assumption of additive separability. That is, the value elicited for the state of interest must not be affected by it being preceded, in Life B, by the lead time in full health. However, this issue is not unique to the lead time approach: additive separability is also an issue with the MVH approach to valuing states worse than dead, where Life A comprises time in poor health followed by time in full health.

Interpretation of the valuation data produced from the lead time TTO as being 'plausible' inevitably relies on expectations about the values for these states reported by other valuation studies, using different elicitation procedures. However, comparisons between mean lead time TTO values and those from other TTO studies, such as the MVH, is not straightforward. This entails either comparisons with valuation data that contain extreme negative numbers, or with artificially bounded ones, neither of which are valid comparisons. Nevertheless, the lead time TTO and the MVH TTO have resulted in comparable proportions of respondents judging a state to be worse than dead in seven out of the 10 EQ-5D states used. Comparisons of the distributions of the *positive* observations across the two studies suggest that the introduction of the lead time may have complex effects on the distribution of overall values. Where there were significant differences between the distributions of positive TTO values between the lead time and MVH in four out of ten states where the mean of the positive TTO values was higher using the lead time approach. While this does not necessarily imply that the overall mean (across positive and negative values) would be different using the lead time TTO, it does suggest further research is required using the approach to value a wider range of states across the EQ-5D descriptive system, to understand what the implications of the lead time approach would be for EQ-5D value sets.

The re-scaled VAS valuations reported here are considerably higher than those reported in the MVH. This may be related to the set of states being considered in our study being principally very poor states of health, in contrast to the MVH study where these states were being evaluated alongside mild states. This potential violation of the independence of irrelevant alternatives has been noted in previous VAS studies (Ling-Hsiang and Kind, 2008).

The effect of background characteristics on valuations suggests some quite complex influences. Those who have themselves experienced severe illness or have experience with poor health as a carer gave *lower* VAS values – whereas these same factors were associated with *higher* lead time TTO values, the latter being consistent with the theory that people with experience of ill health will have a greater understanding of adaptation and will value poor health states more highly. In addition to experience of ill health, higher TTO values were also associated with being in work, having a degree, being a home owner, being a female in a household containing children (suggesting being a mother), and having better self-reported current health on the EQ-VAS. In contrast, the MVH study found only age, sex and employment status to be significant influences on TTO values. One might speculate that these characteristics are associated with a relatively high value placed on time (just ‘being there’) than on health. For example, maybe mothers, workers and those with mortgages to pay value their time more, and are therefore less willing to trade time for health improvements.

As already noted, some aspects of our findings need to be treated with caution. A considerable number of participants regarded the most severe states to be so bad that the 15 year lead time was insufficient to capture their disutility. Our ‘base case’ results are based on the assumption that the minimum value in these instances was -4. This was based on a binary ‘yes/no’ response about whether they would have been willing to trade if the lead time was extended to 20 *and* the duration of H_i was simultaneously reduced to 5. This was used as a pragmatic means of avoiding the complications involved in the use of alternative TTO board front-pieces and maintaining the overall 25 year profile and thus the maximum age of death. The -4 is not a product of iteration toward indifference and did not use any visual aid. It is conservative in the sense that -4 is the value associated with participants being indifferent between 20 years in full health and 5 in *Hi*, whereas their response could be consistent with greater disutility. However, the -4 is not conservative as an estimate of the lead time TTO values that might apply – had we maintained a fixed denominator, but kept extending the lead time, the lead time would need to be 40 years at the point of indifference to yield this same value of -4.

The means of dealing with participants whose dislike of severe states is such that they exhaust the available lead time therefore requires further attention. Altering the duration of H_i is *not* an appropriate means of coping with the issue. As we noted in the Introduction, in relation to the MVH and Torrance protocols, changing both the numerator and the denominator leads to extreme negative values. More fundamentally, a valuation procedure that relies on changing the duration of the state to be valued introduces the problem that the marginal value of time in poor health may not be constant, compounding the problem experienced by all TTOs that the marginal value of time in full health may not be constant (Buckingham and Devlin, 2009). An alternative solution is to extend the lead time. The difficulty here is how the additional lead time will be interpreted by participants: is the entire life profile longer, starting ‘now’ (and therefore *Hi* pushed further into the future than in the TTO task they had previously undertaken, implying expected death in extreme old age for some respondents) or, given they have just done a TTO with a given lead time starting ‘now’, will the *additional* lead time be thought of as extending into the past?

Yet another alternative may be to abandon the 10 year duration for H_i. For example, using a 1-year duration for H_i combined with a 5-year lead time would allow TTO values down to -5, without introducing unrealistically long overall time horizons. Alternative, much shorter durations have been shown to be feasible to value using the lead time TTO approach (Devlin et al 2009). This should reduce the chances of the lead time being exhausted, provided respondents are able to conceptualise different degrees of states worse than dead. However, if at least some respondents are exhausting their lead time, not because this is their genuine quantitative preference, but because they are signalling a qualitative judgement that this is a very poor health state indeed, then an altogether different approach may be called for. Further research is required to design and test alternative ways of handling this issue. The use of electronic props may offer greater flexibility than the physical constraints of a TTO board in this respect.

Our attempts to measure time preference were not successful – the within-respondent inconsistencies suggest these questions were either too demanding or not worded sufficiently clearly. Therefore our results reported here do not control for time preference. Assuming participants employ temporal discounting in TTO exercises, this is a potentially important issue, as the introduction of lead time pushes the state to be valued further into the future, potentially (depending on the durations involved) increasing the effect of time preference on values.

In conclusion, the lead time TTO shows considerable promise as a simple to use, uniform method applicable to all states, which has the potential to avoid the problem in conventional TTO with exaggerated negative values and the requirement arbitrarily to transform those. We are currently undertaking further research to address the remaining issues noted above.

Acknowledgments

This study was funded by the EuroQol Foundation. Ethical consent was obtained from the City University Senate Ethics Committee. An earlier version of this paper was presented at the EuroQol Group meeting (Milan, 2008) and the Health Economists' Study Group meeting (Aberdeen, 2008). We are grateful to Anne Spencer, Queen Mary University of London, and Angela Robinson, University of East Anglia, for their suggestions and advice. We are also grateful to members of the South Yorkshire community who participated in our study. Helpful comments on this paper were received from Katherine Stevens, Leida Lamers and Chris McCabe. The usual disclaimer applies.

References

Buckingham K, Devlin N. (2009) A note on the nature of utility in time and health and implications for cost utility analysis. *Social Science and Medicine* 68:362-367.

Devlin N, Tsuchiya A, Buckingham K, Tilling C (2009). *Does the value of quality of life depend of duration?* Economics Discussion Paper 09/x, City University, London (forthcoming).

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

Drummond MF, Sculpher M, Torrance G, O'Brien B, Stoddart G. (2005) *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press (3rd edition).

Lamers (2007) The transformation of utilities for states worse than death: consequences for the estimation of EQ-5D value sets. *Medical Care* 45(3) 238-244.

Lamers, L.M., McDonnell, J., Stalmeier, P.F.M., Krabbe, P.F.M., Van Busschbach, J (2006) The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Economics* 15(10):1121-1132.

Ling-Hsiang, C., Kind, P. (2008) *Ordinal or cardinal? The VAS strikes back*. Manuscript, www.york.ac.uk/inst/che/images/outcomes/Material_1.pdf

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

Patrick D, Starks H, Cain K, Uhlmann RF, Pearlman R. (1994) Measuring preferences for health states worse than death. *Medical Decision Making* 14: 9-18.

Richardson J, Hawthorne G. (2001) *Negative utility scores and evaluating the AQoL All Worst Health State.* Monash Working Paper 113.

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

Robinson A, Dolan P, Williams A. (1997), Valuing health status using VAS and TTO: what lies behind the numbers? *Social Science and Medicine*, 45, 1289-1297.

Shaw JW, Johnson JA, Coons SJ. (2005) US valuation of the EQ-5D health states: development and testing of the D1 model. *Medical Care.* 43:203-220.

Stalmeier et al (2005) The gap effect: discontinuities in preferences around dead. *Health Economics* 14(7): 679-685.

Szende, A., Oppe, M., Devlin, N. (2007) *EQ-5D valuation sets: an inventory, comparative review and users' guide*. Rotterdam: EuroQol Foundation. Springer.

Tilling C, Devlin N, Tsuchiya A, Buckingham K. (2008) *TTO Valuations of Health States Worse than Dead – A systematic review and framework for analysis*. Health Economics and Decision Science Discussion Paper Series 08/09, School of Health and Related research, University of Sheffield www.shef.ac.uk/scharr/sections/heds/discussion.html

Torrance GW, Thomas W, Sackett D. (1972) A utility maximization model for evaluation of health care programmes. *Health Services Research* 7: 118-33.

Torrance GW. (1982) Multi-Attribute Utility Theory as a Method of Measuring Social Preferences for Health States in Long-Term Care. In *Values and Long Term Care*, pp. 127-56, R. Kane and R. Kane (eds.). Lexington Books Div., D.C. Health Publisher, Lexington, Mass.

Tsuchiya, A., Brazier, B., Roberts, J. (2006), Comparison of valuation methods used to generate the EQ5D and the SF6D value sets in the UK. *Journal of Health Economics*, 25(2): 334-346.

Figure 1. The conventional TTO approach to valuation of states worse than dead.

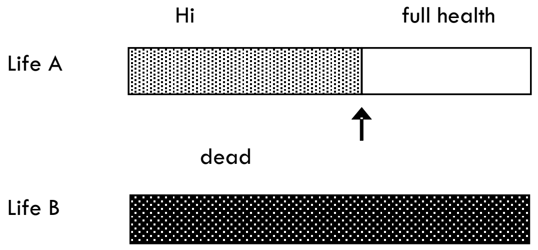
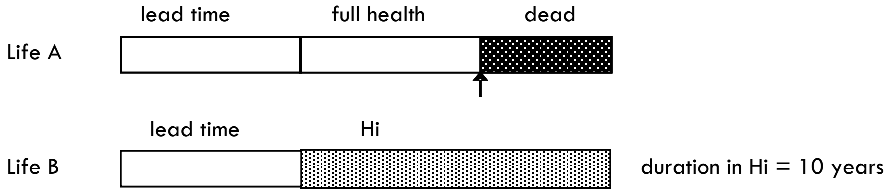


Figure 2. The Lead Time TTO, with an illustration of responses that would yield positive and negative valuations respectively.

State happens to be 'better than dead':



State happens to be 'worse than dead':

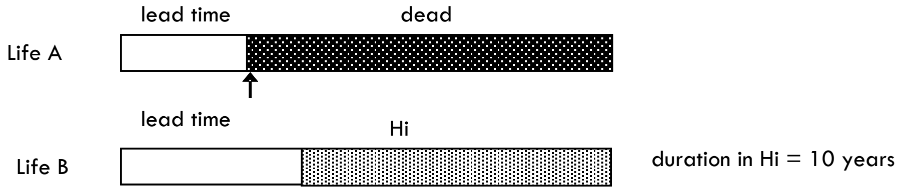
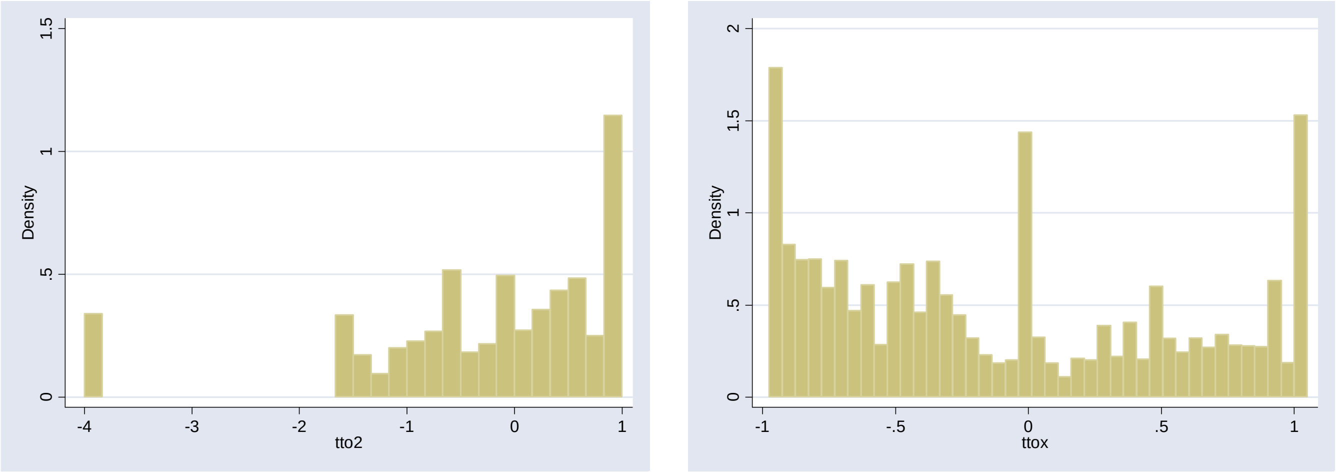


Figure 3. Distribution of TTO results



(a) Current study, all states, untransformed values with minimum value of -4

(b) MVH study, states used in current study, transformed to -1

Table 1. Characteristics of the feasibility study sample

		Feasibility study	MVH sample
Sex:	Female Male	58% 42%	56% 44%
Age: average (standard deviation)	 *<44* *45-64* *65+*	46 (15.4) 45.0% 41.3% 13.7%	47 (18.4) 47.4% 28.9% 23.7%
Net Household income:	Under £10,000 £10,000 to £16,000 £16,001 to £20,000 £20,001 to £26,000 £26,001 to £40,000 £40,001 to £60,000 Over £60,000 No response	14.0% 8.4% 11.2% 4.7% 26.2% 12.2% 5.6% 17.8%	
Education after minimum school leaving age	No Yes	26.9% 73.1%	
Degree or equivalent professional qualification	No Yes	63% 37%	
Main activity	Housework Employment/self-employment Other Retired Seeking work Student	13.9% 55.5% 3.7% 18.5% 4.6% 3.7%	15.5% 49.0% 5.1% 22.5% 5.6% 2.3%
Home ownership status	Own home outright, or with mortgage Rent from local authority Rent from private sector	85.0% 8.4% 6.6%	
Experience of serious illness in you yourself	No Yes	80.6% 19.4%	68.3% 31.7%
Experienced serious illness in your family	No Yes	31.5% 68.5%	28.2% 71.8%
Experienced serious illness in caring for others	No Yes	57.4% 42.6%	72.4% 27.6%
Mean self-reported health on the EQ-VAS		0.81	0.86

Table 2. Raw and re-scaled VAS valuations from the feasibility study, compared with corresponding VAS valuations in the MVH

	Raw VAS, feasibility study			Raw VAS, MVH study			Rescaled VAS, feasibility study*			Rescaled VAS, MVH study*		
	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD
Dead	0	4.61	15.3	0	8.50	15.70	0.0	0.0	*n.a*	0.00	0.00	*n.a*
11111	100	96.29	13.9	100	98.70	4.80	1.0	1.0	*n.a*	1.00	1.00	*n.a*
13332	45	43.14	23.1	20	23.90	16.60	0.43	0.43	0.3	0.16	0.11	0.57
22222	70	66.05	20.2	50	52.30	17.30	0.69	0.66	0.24	0.50	0.45	0.37
32223	30	30.49	19.5	20	22.80	15.50	0.28	0.29	0.20	0.15	0.10	0.56
33323	15	19.79	19.1	10	13.90	12.40	0.10	0.17	0.22	0.07	-0.03	0.04
11112	90	84.18	18.1	87	82.40	15.20	0.86	0.86	0.21	0.87	0.81	0.23
32211	45	42.94	22.9	35	36.30	19.50	0.45	0.42	0.3	0.30	0.28	0.38
33232	20	21.43	17.4	14	16.20	12.70	0.16	0.19	0.2	0.10	0.01	0.71
23232	40	40.47	21.7	25	28.30	16.70	0.39	0.39	0.3	0.21	0.18	0.44
33333	9	13.68	19.1	2	5.60	9.10	0.1	0.11	0.2	0.00	-0.13	0.90
32232	30	30.35	20.1	20	23.40	15.90	0.29	0.29	0.3	0.17	0.06	0.77

* Rescaled values are anchored at 0 = dead and 11111 = 1, using U(Hi) = (Hi – d) / (11111- d). Four participants' data excluded where d > or = 11111

Table 3. TTO Values for Feasibility Study and MVH Study

	Lead Time Feasibility Study – Baseline Restricted to Minimum of -4.0			Feasibility Study with Values Restricted to Minimum of -1.5			Feasibility Study with Values Restricted to Minimum of -1.0 *					MVH Study: Transformed to Minimum of -1.0		
EQ-5D State	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD	LCL (95%)	UCL (95%)	Mean	Median	SD
13332	-0.2	-0.05	1.02	-0.11	-0.05	0.75	0.15	-0.01	0.47	0.06	0.24	-0.23	-0.38	0.55
22222	0.51	0.75	0.78	0.56	0.75	0.55	0.61	0.75	0.44	0.53	0.69	0.5	0.63	0.49
32223	-0.4	-0.15	1.17	-0.23	-0.15	0.74	0.05	-0.04	0.47	-0.04	0.14	-0.19	-0.28	0.56
33323	-0.82	-0.6	1.21	-0.59	-0.6	0.72	-0.13	-0.15	0.42	-0.21	-0.05	-0.39	-0.48	0.49
11112	0.79	0.95	0.36	0.79	0.95	0.36	0.8	0.95	0.31	0.74	0.86	0.82	0.93	0.29
32211	0.06	0.35	1.1	0.17	0.35	0.74	0.32	0.35	0.52	0.22	0.42	0.14	0.25	0.6
33232	-0.66	-0.55	1.28	-0.43	-0.55	0.78	-0.05	-0.14	0.47	-0.14	0.04	-0.33	-0.43	0.51
23232	0.05	0.23	0.81	0.07	0.23	0.72	0.26	0.23	0.46	0.17	0.35	-0.1	-0.08	0.59
33333	-1.16	-0.95	1.34	-0.79	-0.95	0.68	-0.25	-0.24	0.41	-0.33	-0.17	-0.54	-0.65	0.41
32232	-0.51	-0.47	1.11	-0.37	-0.47	0.76	0	-0.12	0.44	-0.08	0.08	-0.23	-0.38	0.57

* the feasibility study results were rescaled to achieve a minimum value of -1.0 by dividing all the negative scores of the baseline analysis by 4 (their maximum value).

Table 4. Further analysis of the characteristics of TTO valuation data for SWD and SBD, compared to the MVH

	Lead time TTO feasibility study (n=109)						**MVH study**							**Two sample Kolmogorov-Smirnov test of distributions$_b$**
	Baseline value at -4	**TTO values worse than dead**	**Distribution of TTO values better than dead**				**Observations per state**	**TTO values worse than dead**		**Distribution of TTO values better than dead**				
EQ-5D State	%	%	Mean	25 percentile	Median	75 percentile	n	%	chi^2 test p-value[a]	Mean	25 percentile	Median	75 percentile	p-value
13332	4	51	0.57	0.35	0.55	0.79	737	63	**0.03**	0.46	0.23	0.40	0.68	**0.02**
22222	2	15	0.78	0.56	0.89	0.95	770	13	0.50	0.67	0.50	0.68	0.93	**0.00**
32223	7	55	0.49	0.26	0.48	0.65	749	59	0.40	0.48	0.23	0.48	0.63	0.24
33323	10	74	0.40	0.18	0.30	0.55	761	77	0.49	0.43	0.20	0.40	0.58	0.34
11112	0	6	0.87	0.85	0.95	1.00	1207	2	**0.01**	0.86	0.80	0.95	1.00	**0.01**
32211	5	32	0.62	0.40	0.65	0.95	745	34	0.72	0.56	0.33	0.53	0.80	0.10
33232	10	63	0.45	0.19	0.40	0.66	757	71	0.09	0.44	0.28	0.40	0.63	0.72
23232	1	37	0.60	0.35	0.55	0.95	726	51	**0.01**	0.50	0.30	0.50	0.73	**0.01**
33333	16	83	0.38	0.15	0.30	0.56	2997	86	0.42	0.43	0.13	0.38	0.68	0.64
32232	6	66	0.49	0.30	0.45	0.70	749	64	0.77	0.51	0.30	0.48	0.73	0.61

a: Null hypothesis: the proportion of negative observations is the same across the two studies.
b: Null hypothesis: the cumulative distribution of the positive observations is the same across the two studies.