Time to tweak the TTO. But how?

MM Versteegh and AE Attema and M Oppe and NJ Devlin and EA Stolk

Erasmus University Rotterdam

15. March 2012

Time to tweak the TTO. But how?

Matthijs M. Versteegh[a] *
Arthur E. Attema[a]
Mark Oppe[a]
Nancy Devlin[b]
Elly A. Stolk[a]

[a] iMTA/iBMG, Erasmus University of Rotterdam, Institute of Health Policy & Management / Institute for Medical Technology Assessment, PO Box 1738, 3000 DR Rotterdam, The Netherlands.
[b] Office of Health Economics, London UK.

* Correspondence to: Matthijs Versteegh
E-mail: versteegh@bmg.eur.nl
Phone: +31 10 4089170

Word count: 5.003

This study was funded by ZonMW: the Netherlands organization for health research and development [project number 152002039] and the EuroQoL foundation.

Abstract:

This paper examines the effect different specifications of the Time Tradeoff (TTO) task have on health state values. The new lead time TTO is compared to an equally viable method called lag time TTO, both in two time frames. We test whether the two methods yield comparable health state values and whether the relative importance of dimensions of health is similarly stable between TTO specifications. The tasks were applied online and compared to results from a study with identical TTO specifications but with a different mode of administration. Lag time TTO produced lower values than lead time TTO and the difference was larger in the longer time frame. The relative importance of different dimensions of health was affected by the duration of the health state. Generally, the lead time TTO performed in group sessions gave more favorable results than the online exercise based on feasibility and data quality.

1. Introduction

Adequately measuring health state values is of great importance to the allocation of resources in health care. Health state values are used to calculate Quality Adjusted Life Years, which is a metric for the effect of health care interventions. Attempts to improve the measurement of health state values have led to several methodological innovations. First, novel specifications of the original Time Tradeoff (TTO) method (Torrance et al. 1972) have been developed to improve the measurement of health states considered 'worse than dead'. One novel specification, the lead time TTO (Robinson, Spencer 2006), has been proven feasible (Devlin et al. 2011). Alternatives to the lead time TTO, such as lag time TTO, have been suggested and are in principle equally capable of addressing issues in the valuation of health states worse than dead. However, there is relatively little evidence on how these methods compare and there may be complex differences between them (Devlin et al. 2010). Second, researchers have explored the internet as a tool for administering the TTO task to have more respondents at lower costs (Norman et al. 2010). In this study, lead time TTO and lag time TTO are compared in a between-subject design, and results from an online setting are compared to results from a study in which the TTO was self-completed with interviewer guidance provided by two interviewers to a group of respondents.

Valuation methods are used to determine the desirability of a hypothetical state of health through assigning a value. In the TTO, a value is assigned by letting respondents trade off length of life against quality of life. The resulting value is generally taken to reflect the health-related quality of life per period an individual enjoys for the duration of that health state. The value is elicited through asking respondents if they would prefer living x years in a period of full health to living t years in impaired health where $x < t$. If respondents accept living a shorter period *t* in full health, they are essentially willing to trade length of life for quality of life. The health state value is then given by x/t. When respondents indicate they would rather trade off all healthy life years than having to live in a particular health state for period *t*, they indicate that this health state is worse than dead, at least when the duration of that health state is equal to period *t*. Respondents then enter a different task to measure their negative preference values (since $x < 0$). In this 'worse than dead' task, respondents are asked to choose between immediate death and a life of duration *t*, with *x* years in full health preceded by t-x years in the imperfect health state. The value for the health state following this 'worse than dead' task is $-x/(t-x)$. The 'classic' TTO has been criticized for having different valuation procedures to elicit values for health states better and worse than dead. Due to the use of two procedures, TTO values may not 'lie on the same underlying utility scale' (Tilling et al. 2010). Furthermore, sacrificing one additional year in the worse than dead procedure leads to a non-linear marginal decrement in the value of a health state (Attema et al. 2012), and while values for health states better than dead are restricted between 0 and 1, health state values measured with the procedure for worse than dead can become very low (Devlin et al. 2011, Lamers 2007), which subsequently requires an arbitrary transformation of those values.

Alternative specifications of TTO that would overcome the above mentioned problems would thus apply one method for both worse than dead and better than dead health states, would have an iteration procedure generating equal changes in value for each step, and would avoid an arbitrary transformation of the TTO values. Two alternative specifications of TTO that would meet these requirements are the lead time TTO and the lag time TTO. The lead time TTO was first proposed by Robinson and Spencer (Robinson and Spencer 2006). In this TTO specification, extensively discussed elsewhere

(Devlin et al. 2011), the impaired health state 'begins' after a period of healthy years (the lead time), rather than immediately. The methods provided consistent results in 159 undergraduate students (Attema et al. 2012) as well as in 109 members of the general population (Devlin et al. 2011),and showed that total time frame and ratio of lead time to disease time influenced health state values. We are only aware of one study testing lag time TTO (Devlin et al. 2010). In lag time TTO, healthy life years *follow* the impaired health state, rather than *preceding* it. Although the lag time TTO would equally tackle the above mentioned problems of 'classic' TTO, it did not produce the same values as lead time TTO in a study using 7 EQ-5D health states (Devlin et al. 2010). In this study, which used 5 years disease time and 10 years lead/lag time, lead time TTO values were lower for more severe states than lag time values. However, in lag time TTO more people were willing to trade off time for mild states, be it less on average (i.e. higher mean values) than in lead time TTO. Thus, the findings were mixed regarding the specific effect of the specification of TTO on health state values. In lead time TTO, the health state under valuation is further away in the future than in lag time TTO, where the health state 'begins' immediately. It could be hypothesized, therefore, that lead-time values for the same health state will be higher than lag time values if respondents have positive time preferences, which is frequently observed (Olsen 1994, Gyrd-Hansen 2002), although there are also reports of negative time preferences for TTO (Dolan and Gudex 1995). Alternatively, it could be hypothesized that lag time TTO results in higher values, since the lag-time of full health after the health state might be interpreted as being cured from the health state, which, arguably, influences the perception of the severity of the health state. Conceptually, lag time TTO might be more 'plausible' for mild states and curative treatments, since poor health is followed by good health. Lead time TTO may be more plausible for very severe health states and preventive treatments since the health state starts in the future and is followed by death. In order to provide guidance for the preferred TTO specification, lag time TTO additional comparisons to lead time TTO, which is the main focus of this study.

Differences in health state values are only attributable to the specification of the TTO if all other elements of the task are equal. Indeed, many elements in a health state valuation task can influence health state values (Stalmeier et al. 2001). Likewise, in order to attribute differences in data quality and study feasibility to the mode of administration, it is crucial that other task elements are equal. Valuation tasks have been administered in many different settings such as face to face interviews at the respondent's home (Brazier et al. 2002), self completed by individuals with interviewer guidance available and with multiple respondents in one room (Stolk and Busschbach 2003, Versteegh et al. 2012), via postal questionnaires (Devlin et al. 2003) or via the internet, which is known to produce lower data quality for 'classic' TTO (Norman et al. 2010), but may facilitate a good geographical coverage of respondents at low costs(Bansback et al. 2012). In the current study, data quality and feasibility of administering a TTO task over the internet is compared to data quality and feasibility of a TTO which was self completed by participants with interviewers available to multiple respondents in one room (from now on referred to as 'group TTO'). An extended 'checklist' is provided to indicate the comparability of other elements of the TTO task in these two settings.

The paper is organized as follows. First, we introduce the study design, the different TTO variants applied, and the analyses. Second, the result section compares lead time TTO to lag time TTO, and compares data quality and feasibility of the online setting to those of the group setting. Last, we discuss and interpret our findings.

2. Methods

2.1 Respondents

Respondents were sampled from members of a commercial panel and stratified to represent the Dutch population based on gender and education. Respondents were also stratified to match the Dutch population for age, but only respondents between 18 and 65 were approached to participate in the online experiment. Respondents did not receive a direct financial reward for participating.

2.2 Health state selection and description

Health states were based on the Dutch version of the EQ-5D 5-level (EQ-5D-5L). This instrument consists of 5 dimensions of health: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The instrument has 5 answer categories for each dimension, generating 3125 (5^5) health states. Of the total amount of possible health states, 100 were selected based on a previously developed D-optimal design (Oppe and Van Hout 2009).

2.3 Study design

Respondents were asked to perform a combination of tasks. They first filled out background questions and indicated their own health on the EQ-5D-5L instrument and the EQ-5D visual analogue scale ranging from 0 to 100 where 0 represented worst imaginable health and 100 represented best imaginable health.. After these initial questions, respondents were randomized over two arms with two different choice based tasks. The first arm consisted of a best-worst scaling task, where respondents had to indicate the best attribute level combination and the worst attribute level combination of EQ-5D-5L health states. In the second arm, respondents had to choose which of two EQ-5D-5L health states they considered best in a paired comparison task. After this, respondents were randomized over 5 TTO tasks. Within the 5 TTO tasks, respondents were randomized over 10 blocks containing 10 EQ-5D-5L health states, and each health state was presented in random order. The study ended with several questions about the feasibility of the TTO tasks, discussed later on.

2.3.1 The TTO tasks

The TTO tasks were preceded by an animated instruction. In the animation, it was explained to respondents how to trade off life years using an example with a hypothetical EQ-5D state and an animated 'doctor' who pointed out the several elements of the task. The animation was specifically designed to reflect the characteristics of the different TTO tasks. Thus, the TTO examples in each animation preceding the real TTO task were identical in characteristics and lay-out as the real TTO task that followed. A schematic description of the five TTO exercises described in this study and the corresponding utility equations are presented in figure 1. The utility equations in figure 1 are discussed in detail later on.

[FIGURE 1 ABOUT HERE]

As can be seen from figure 1, the classic TTO is a two-part task with different visual representations and different utility equations for health states better than dead (BTD) and health states worse than dead (WTD). The other 4 TTO tasks have a uniform visual representation and utility equation for BTD and WTD valuations. In all tasks, respondents were asked to choose between a fixed period in life A and a variable period *t* in life B. The value of *t* was dependent on the previous choice for either life A or B by respondents and followed a fixed iteration procedure described below.

2.3.2 Iteration procedure

The iteration procedure followed that of the Measurement and Valuation of Health protocol (MVH) and was adapted for different ranges of *x*. The first two 'steps' of the fixed iteration procedure were similar for the different TTO tasks. At the first iteration, respondents were asked to choose between living in life A, which contained the health state and depending on the task a lead time or lag time in full health, or life B, which was set at the maximum of all years in full health (utility=1, or *x*=10, 15 or 20, depending on total time frame). At the second iteration, life B had a value of utility=0 (or *x*=0 for the classic TTO and *x*=10 for the other variants). If respondents favor life B at utility=0, they indicate that the health state is worse than death. If they favor life A, they indicate that the health state is better than death. After this 'sorting question', the iteration procedure continued with a choice between life A and life B where B has value *x* for utility=0.5 or -0.5. Conditional on choosing life A, or B, the remaining iterations represented utility increments or decrements of 0.1 or 0.05 with the corresponding values of *x* in life B.

2.3.3 Utility equations

To define and clarify the equations from figure 1, an example of the lead time TTO and lag time TTO in a 20 year time frame is shortly discussed below. Assuming no discounting, the utility equation in lead time TTO in a 20 year time frame is:

1) $$10U_{FH} + 10U_{HS_i} = xU_{FH}$$

where U_{FH} is the utility of full health, U_{HSi} is the utility value of the health state *i* and *x* is the number of years in full health at which the respondent indicated being indifferent in the TTO task. Solving for U_{HSi} gives:

2) $$U_{HS_i} = \frac{x-10}{10}$$

For a respondent who considers *x*=13 years in full health equal to 10 years in U_{HSi} the utility value for U_{HSi} = (13-10)/10 = 0.3. Equally, for lag time TTO the utility equation is:

3) $$10U_{HS_i} + 10U_{FH} = xU_{FH}$$

Equation 3 can also be solved for U_{HSi}, which again results in equation 2. Hence, the QALY model with no discounting predicts equal answers for lead and lag time TTO.

2.4 Comparator study

To inspect the effect of using an internet panel, results were compared to a pilot study with Dutch respondents (N=201), which used the same digital aid that was used by the internet panel, with interviewers present. The digital aid was an exploratory version of the EuroQoL Valuation Technology (EQ-VT ©2011, EuroQoL Group), developed for use in a pilot study involving multiple countries. The comparator study used the same 100 health states but was conducted in a group setting with a plenary introduction, and thus, arguably, in a better controlled setting. The Dutch group TTO study employed a lead time TTO study with a 10 year lead time and 5 year disease time, which is similar to the lead time study with a 15 year time frame in figure 1. The two TTO studies were nearly identical except for the mode of administration. TTO in a group session has been shown to be feasible in the general public for classical TTO (Stolk and Busschbach 2003), but, to date, has only been successfully applied in university students for lead time TTO (Attema et al. 2012). As TTO studies can differ on very many aspects, with varying influence on outcomes, we developed a TTO comparison checklist (table I), partly based on an article discussing these issues (Stalmeier et al. 2001). This checklist is of importance to this study, since it shows that the specification of lead time TTO in the online setting was nearly identical to the specification in the group setting. Therefore, differences in data quality were attributable to differences in the setting of administration rather than due to other types of study heterogeneity.

[TABLE I ABOUT HERE]

2.5 Analyses

2.5.1 Lead time TTO vs. Lag time TTO

Mean lead time TTO and lag time TTO values are compared for all 100 health states. Different ranges of attainable utility values distort comparisons of the mean between tasks. For example, solving the utility equations from figure I for t=0 (trading in all life years) results in U=-2 for a 15 year time frame and U=-1 for a 20 year time frame. Therefore, comparisons of the mean are only made for tasks with similar attainable utility values. Also, utility values produced by the different tasks cannot be compared to a non-experimental EQ-5D-5L tariff, as the valuation of the EQ-5D-5L is currently still under development. To get an indication of the convergent validity of the values produced in the online exercise, they were compared to estimated EQ-5D-5L values based on a mapping function (van Hout et al. In press). These estimated values represent which utility value is expected for an EQ-5D-5L state based on previous valuations for the EQ-5D-3L.

Probably the most important application of either lead time TTO or lag time TTO is the valuation of health state descriptive systems such as EQ-5D. These descriptive systems often consist of multiple dimensions of health. The relative importance of the dimensions of EQ-5D in the different specifications TTO is compared through random effects regression analysis to take account of the panel structure of the data (multiple TTO observations per respondent). Although the sizes of the coefficients are not directly comparable due to different ranges of the dependent variable (the TTO score), the relative importance of dimensions within each regression model can still be compared. Predictor variables were the EQ-5D dimensions of health as continuous variables.

2.5.2 Data quality

Several criteria were used to assess the quality of the data produced by the different TTO tasks. Respondent agreement in the different TTO tasks was addressed by comparing variances with Levene's test and Brown & Forsythe tests. The assumption here was that differences in valuations between respondents, regardless of what causes these differences, results in increased variance and thus less precise utility estimates. Although larger standard deviations may reflect preference heterogeneity rather than poorer data quality, a valuation method is arguably more preferable if there is more agreement amongst respondents. Variances for classic TTO (with transformed negative values) were only compared to the TTO tasks with a 20 year time frame as TTO values for these two lie on the same -1 to 1 scale, rather than the TTO values of the TTO tasks with a 15 year time frame, which lie on a -2 to 1 scale and, thus, logically have larger variances. Standard deviations, which have a more intuitive interpretation than variances, are plotted for lead time TTO and lag time TTO. Other indicators of data quality were whether respondents were willing to trade off any time at all (non-traders), how many iterations respondents used before reaching their point of indifference,, how many respondents 'used up' all tradable time and how many respondents did not differentiate between health states.

2.5.3 Feasibility

Differences between tasks were compared using four items of a feasibility questionnaire presented after the TTO task. The items tested whether the instructions were clear, if the questions asked were easy, if it was difficult to reach the point of indifference and if it was easy to tell the difference between the different health states under valuation. Answer categories ranged from 1 (completely agree) to 5 (completely disagree). Mean scores of the different tasks were compared to each other as well as to mean scores in the group TTO.

Since respondents valued multiple health states (5 in the group TTO and 10 in the online study), and data quality is known to be affected by learning effects (Augestad et al. 2012), we repeated the analysis using only the first 5 valued health states in the online study. We tested for significance of order effects through regressing the sequence number of a health state on the amount of iterations using OLS, as proposed by Augestad et al. (2012).

2.6 Exclusion criteria

All respondents who completed the online exercise were included in the analyses. Analyses were rerun in a smaller sample without respondents who: 1) indicated on the feasibility questionnaire they did not understand the task 2) did not differentiate between any of the 10 health states and 3) had used only 3 or fewer iterations for all health states to check for consistency of findings.

2.7 Software

Statistical analyses were run in STATA 11.

3. Results

6222 respondents finished all of the online tasks. The resulting dataset was a balanced panel with 10 TTO observations for each respondent. Respondents in the online panel were slightly older than the Dutch population average (*mean*=42.3 (*sd*=14.2) vs. Dutch population mean of 2009 = 40.1) and contained more females with 58.3 percent female and 41.7 percent male, compared to a nearly 50/50 distribution in the Netherlands. Mean self-assessed health on the EQ-5D Visual Analogue Scale (VAS) for the online population was 76.7 (*sd*=17.4), which compares to a mean VAS of 83.3 (*sd*=13.6) in the group TTO that contained 51.2% females and had a mean age of 42.1 (*sd*=14.1). OLS regression indicated that respondents used fewer iterations ($p<.001$) for health states presented later in the sequence, on average 0.4 iterations less than the previous health state for each consecutive health state. Therefore, where relevant, results were rerun using only the first 5 health states of the online study to allow better comparability with the group TTO.

3.1 Lead time TTO vs. Lag time TTO

Lead time TTO resulted in systematically higher values than lag time TTO for the 20 year time frame (on average 0.25 higher) with larger average differences for poorer health states (figures 2a & 2b). In the 20 year time frame, none of the lag time values were higher than the lead time values. Results for the 15 year time frame were mixed: on average lead time TTO values were 0.13 higher in the 15 year time frame and lower than lag time TTO values for 18 out of 100 health states (28 out of 100 using first five health states). The range of utility values in the 15 year time frame was 1.13 for lead time TTO (from -0.4 to 0.73) and 1.14 for lag time TTO (-0.46 to 0.68). In the 20 year time frame, values where higher than in the 15 year time frame for both variants, which is most likely due to the range of attainable values in the 20 year time frame (the minimum value of the 15 year time frame was -2, compared to -1 in the 20 year time frame. The minimum value of -1 also influenced the observed range of values in the 20 year time frame, which was smaller for both variants with a range of 0.69 for lead time TTO (0.20 to 0.89) and 0.80 for lag time TTO (-0.08 to 0.72). As can be seen from figures 2a and 2b, the range of values produced by the lead time TTO and the lag time TTO was smaller than would be expected based on the estimated EQ-5D-5L values from the literature (van Hout et al. In press). In a previous Dutch valuation study of EQ-5D-3L, using 'classic' TTO, the worst health state (33333) value was -0.39 and the second best health state (11211) was 0.897, a range not reflected in any of the TTO specifications tested here (Lamers et al. 2006). Excluding respondents that claimed not to understand the task, respondents that did not differentiate between health states or used less than 3 iterations, did not alter this finding. Similarly, the utility values of the classic TTO, with a transformation for negative values to be bound at -1 as applied in the previous TTO valuation studies of EQ-5D-3L, did not produce negative mean values for any of the health states and thus also had a rather limited range of values compared to previous EQ-5D valuation studies (Lamers et al. 2006, Dolan 1997).

[FIGURE 2A / 2B ABOUT HERE]

The specification of the TTO task influenced the relative importance of the different dimensions of health (table II). The size of the coefficients represents the marginal decrement in utility caused by scoring one point higher in a particular dimension on the five level descriptive system. The order of the relative importance of different dimensions of health was not affected by the choice for lead time TTO or lag time TTO, but by the duration of the health state. In the 20 year time frames, with a disease duration of 10 years, the health dimensions 'Anxiety/Depression' was considered worse than 'Pain/Discomfort' while the inverse was found for the 15 year time frame, which has a disease duration of 5 years. Equally, problems in usual activities were considered more problematic than problems with self-care in the 20 year time frame while the inverse was found for the 15 year time frame. The order in the 'classic' TTO was different from the order in the lead time TTO and lag time TTO. Furthermore, the relative importance of dimensions of health in the group TTO was different from those in the online study, despite the similar specification of the TTO task. The regression models using only the first 5 health states from the online study gave identical orderings as found using all 10 health states.

[TABLE II ABOUT HERE]

3.2 Data quality and response characteristics

Figure 3 shows the standard deviations for the lead and the lag time TTO in both time frames. Lag time TTO tasks had a larger variance than lead time TTO for nearly all health states. Both Levene's test and Brown & Forsythe test suggest that the mean variance of lag time TTO is indeed higher in both the 15 year time frame ($p<.001$) and the 20 year time frame ($p<.001$). Using the same test statistics, the classic TTO with transformed negative values has a smaller variance than lag time TTO ($p<.001$), but larger variance than lead time TTO ($p<.001$). When only respondents were included who had indicated on the feasibility questionnaire that they thought the task was clear (answer 1 on question 1), that they understood the task (answer 1 on question 2) and had not valued all 10 health states equal, all statistical tests gave significant differences ($p<.001$). The mean standard deviation (averaged over all health states) of the group TTO was 0.65 (N=201), which compares to the mean standard deviation of 0.81 (N=1067) of the online lead time TTO in a 15 year time frame. When only respondents were included that were randomized to the LT-TTO in a 15 year time frame, and indicated they thought the task was clear and understood the task, the mean standard deviation increased somewhat to 0.83 (N=359). Using only the first 5 valued health states from the online study increase the mean standard deviation of the lead time TTO in a 15 year time frame to 0.84 (N=533).

[FIGURE 3 ABOUT HERE]

The number of non-traders (%Utility = 1) and the distribution of BTD (Utility > 0) and WTD (Utility < 0) responses in the online task suggest that the lead time TTO causes respondents to judge health states as being less severe compared to lag time TTO (table III). Also, the online panel had different response characteristics than the group TTO. In the online study, each TTO task had more respondents who did not differentiate between the 10 health states they were asked to value, more respondents who used only a couple of iterations to indicate indifference, more non-traders (utility =1) and fewer states were valued as worse than dead. Interestingly, the group TTO showed also showed a large percentage of respondents valuing a state equal to being dead (utility=0). Using only the values of the first 5 health states from the online study had similar results. For example, still more than 60% of respondents used only 4 or less iterations and about 35% of the sample valued health states at utility = 1.

[TABLE III ABOUT HERE]

3.3 Feasibility

The apparent differences between group lead time TTO characteristics and characteristics of the TTO specifications in the online task may be attributable to the feasibility of an online TTO. Indeed, respondents in the online panel considered the task much less clear and more difficult to understand than respondents in the group TTO when this hypothesis was tested with a t-test ($p<.001$) (table IV).

[TABLE IV ABOUT HERE]

Respondents who disagreed with statement 1 and 4 were generally older than the population average in both the online sample and the group TTO. There were no clear patterns between feasibility statements and gender or health of the respondents as measured by VAS.

4. Discussion

In this study novel specifications of the TTO were compared to explore the impact of alternative specifications on health state values. Results from administering TTO tasks through the internet were compared to results from a group setting.

The specifications of the TTO tasks applied in this study systematically affected health state utilities and the relative importance of dimensions of health. In the 20 year time frame, lag time TTO produced lower values than lead time TTO with mixed results for the 15 year time frame. Interestingly, the relative importance of different dimensions of health was affected by the duration of the impaired health state, but not by the choice for lead time TTO or lag time TTO. Apparently, respondents considered Anxiety/Depression to be worse than Pain/Discomfort only for a duration longer than 5 years. We also found that the lead time TTO performed by respondents with interviewer assistance available gave more favorable results than the online exercise, based on feasibility and data quality.

Lag time TTO were expected to produce lower values than lead time TTO, due to positive time preferences. It was, however, equally possible for lag time TTO to produce higher values due to (less frequently observed) negative time preferences, or due to the fact that in lag time TTO the health state may seem less problematic as full health returns when the imperfect health state ends. On average, the effect of time preference (i.e., preferring to be in the best health state immediately) on health state values is larger than the 'preference for improvement' effect. From these findings, it seems that the additive separability assumption of the QALY model (i.e. a health state value is independent of health states preceding or following it) does not hold, since here it is shown that health state utilities elicited with lag time TTO are lower than lead time TTO in the 20 year time frame. A 1995 study into time preferences and the duration of health states by Dolan and Gudex (1995) compared lead time TTO with lag time TTO, be it without using those exact terms for the TTO specifications. That study had a lead time TTO and a lag time TTO with nine years in full health and one year in an impaired health state. For three out of five health states lead time median values were lower than lag time values. Thus, for three out of five health states respondents considered having the health impairment earlier in time preferable to having the health impairment later in time (i.e. negative time preferences). Although this finding seemingly contradicts the results presented here, it may equally well be that individuals yield more utility out of having the health impairment earlier in time when the duration of the health state is relatively short, for example, to get the health state 'over with'. This would be in line with our finding that for the shorter disease duration the difference between lead time TTO and lag time TTO is smaller. These results highlight the influence of time preference in TTO tasks, especially when the addition of lead or lag time increases the considered time horizon. A detailed study into correcting the TTO values from this study for time preferences is currently underway.

The relative importance of different dimensions was affected by the duration of the health state in the experiment. Although all different variants tested indicated that the dimensions 'Pain/Discomfort' and 'Anxiety/Depression' caused the largest decrement in health state utilities, the 'Anxiety/Depression' dimension received larger weight at longer durations in all three TTO tasks. If the relative importance of an attribute of a health state depends on duration, it is unlikely that the specific utility decrement can be extrapolated to durations other than the one applied in the TTO task. Interestingly, the 15 year lead time TTO in the group setting had a different ordering of the dimensions 'Mobility' and 'Self-care' than the 15 year lead time TTO in the online study, which suggests that the importance of attributes within lead and lag time TTO is also influenced by the mode of administration, as has been observed earlier for 'classic' TTO.

Although instructions for the online TTO were very carefully designed by a team of researchers with experience in TTO and consisted of both textual and graphical explanations of the task, the online TTO experiment performed poorer in terms of feasibility and data quality than the group TTO. Task engagement was low in the online setting. Roughly two thirds of observations used maximally four iterations to determine the time in good health that could be traded off to avoid being in the impaired health state. With the iteration procedure used in this study, this means that two thirds of the health states were valued at either 1 (1 iteration), 0 (2 iterations), 0.5/-0.5 (3 iterations) or 0.6/-0.6/0.4/-0.4 (4 iterations). Although it is possible that respondents did not know their preference more precisely than represented by one of these utility values, comparison with the group TTO suggests that precision may be improved in a different setting. TTO data for health states derived from questionnaires is generally used to estimate prediction models which estimate utility values for all possible health states, based on the health states used in the TTO study. Increased variance relative to other TTO methods is likely to negatively affect these prediction models.

The lowered data quality, compared to the group TTO, resulted in very large standard deviations for mean values, or, in other words, much heterogeneity in the data. This heterogeneity was largest for lag time TTO variants, suggesting that respondents differ more in their answers in this task than in classic TTO or lead time TTO, which could be due to several unknown variables. Although these results seem to indicate that respondents were better able to grasp the lead time TTO task, leading to less difference in answers, such a conclusion would not fully align with the self-reported feasibility of the task. The latter indicates that lead time was, on average, considered easier than lag time TTO only in the 15 year time frames, whilst the difference was the other way around for the 20 year time frames. The increased variance in the lag time TTO tasks is thus not solely attributable to understanding of the task.

The many utility = 0 valuations in the Dutch group TTO setting might reflect a misunderstanding of the lead time TTO by respondents, since the effect was observed both in the group TTO and the online tasks. Indeed, the data suggests, both for group and online modes of administration, there is an incentive to 'short-cut' the task and complete it quickly rather than thoroughly. Interviewer presence seems to improve results, but perhaps face-to-face interviews might be the preferred mode of administration for TTO tasks involving iteration procedures. Task understanding seems to plays a role. In lead time TTO in a 15 year time frame, a health state is valued at '0' if a respondent is indifferent between living 10 healthy years, followed by 5 years of impaired health in life A and living 10 healthy years in life B. If respondents do not understand that in life B they trade in 5 healthy years, rather than 5 years in impaired health, the two states may seem equally valuable, causing the respondent to indicate 'indifference'. Further research is required to see if our results are replicated in face-to-face interviews.

5. Conclusion

Lead time TTO and lag time TTO seem equally feasible, but yield different health state values. Differences between lead time TTO and lag time TTO are systematic and may well be attributable to time preferences of respondents, which requires further study. Given the differences in findings between the online study and the group TTO study, it seems that interviewer presence greatly improves the quality of health state values.

Acknowledgements:

We would like to thank Koonal Shah sharing his thoughts on the lead time TTO and for his contributions to the TTO checklist.

Conflict of interests:

Authors declare to have no conflicts of interests.

References

Attema, A.E., Versteegh, M.M., Oppe, M., Brouwer, W.B.F. & Stolk, E.A. 2012, "LEAD TIME TTO: LEADING TO BETTER HEALTH STATE VALUATIONS?", *Health Economics,* , 2012 Online early view.

Augestad, L.A., Rand-Hendriksen, K., Kristiansen, I.S. & Stavem, K. "Learning Effects in Time Trade-Off Based Valuation of EQ-5D Health States", *Value in Health,* , 2012 Online early view.

Bansback, N., Tsuchiya, A., Brazier, J. & Anis, A. 2012, "Canadian Valuation of EQ-5D Health States: Preliminary Value Set and Considerations for Future Valuation Studies", *PloS one,* vol. 7, no. 2, pp. e31115.

Brazier, J., Roberts, J. & Deverill, M. 2002, "The estimation of a preference based measure of health from the SF-36", *Journal of health economics,* vol. 21, pp. 271-92.

Devlin, N.J., Buckingham, K., Shah, K., Tsuchiya, A., Tilling, C., Wilkinson, G. & van Hout, B. 2010, *A comparison of alternative variants of the lead and lag time TTO.* OHE Research Paper London: Office of Health Economics.

Devlin, N.J., Hansen, P., Kind, P. & Williams, A. 2003, "Logical inconsistencies in survey respondents' health state valuations -- a methodological challenge for estimating social tariffs", *Health Economics,* vol. 12, no. 7, pp. 529-544.

Devlin, N.J., Tsuchiya, A., Buckingham, K. & Tilling, C. 2011, "A uniform time trade off method for states better and worse than dead: feasibility study of the 'lead time' approach", *Health Economics,* vol. 20, no. 3, pp. 348-361.

Dolan, P. 1997, "Modeling valuations for the EuroQol health states", *Medical care,* vol. 35, pp. 1095-1108.

Dolan, P. & Gudex, C. 1995, "Time preference, duration and health state valuations", *Health Economics,* vol. 4, no. 4, pp. 289-299.

Gyrd-Hansen, D. 2002, "Comparing the results of applying different methods of eliciting time preferences for health", *The European journal of health economics : HEPAC : health economics in prevention and care,* vol. 3, no. 1, pp. 10-16.

Lamers, L.M. 2007, "The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets", *Med Care,* vol. 45, no. 3, pp. 238-44.

Lamers, L.M., McDonnell, J., Stalmeier, P.F.M., Krabbe, P.F.M. & Busschbach, J.J.V. 2006, "The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies", *Health Econ,* vol. 15, no. 10, pp. 1121-32.

Norman, R., King, M.T., Clarke, D., Viney, R., Cronin, P. & Street, D. 2010, "Does mode of administration matter? Comparison of online and face-to-face administration of a time trade-off task", *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation,* vol. 19, no. 4, pp. 499-508.

Olsen, J.A. 1994, "Persons vs years: two ways of eliciting implicit weights", *Health economics,* vol. 3, no. 1, pp. 39-46.

Oppe, M. & Van Hout, B.A. 2009, "The optimal hybrid: Experimental design and modeling of a combination of TTO and DCE", *27th Scientific Plenary Meeting of the EuroQol Group - Proceedings*, ed. J. Yfantopoulos, , pp. 61-72.

Robinson, A. & Spencer, A. 2006, "Exploring challenges to TTO utilities: valuing states worse than dead", *Health Econ,* vol. 15, no. 4, pp. 393-402.

Stalmeier, P.F., Goldstein, M.K., Holmes, A.M., Lenert, L., Miyamoto, J., Stiggelbout, A.M., Torrance, G.W. & Tsevat, J. 2001, "What should be reported in a methods section on utility assessment?", *Medical decision making : an international journal of the Society for Medical Decision Making,* vol. 21, no. 3, pp. 200-207.

Stolk, E.A. & Busschbach, J.J.V. 2003, "Validity and feasibility of the use of condition-specific outcome measures in economic evaluation", *Quality of Life Research,* vol. 12, no. 4, pp. 363-71.

Tilling, C., Devlin, N., Tsuchiya, A. & Buckingham, K. 2010, "Protocols for time tradeoff valuations of health states worse than dead: a literature review", *Medical decision making : an international journal of the Society for Medical Decision Making,* vol. 30, no. 5, pp. 610-619.

Torrance, G.W., Thomas, W.H. & Sackett, D.L. 1972, "A utility maximization model for evaluation of health care programs", *Health Serv Res,* vol. 7, no. 2, pp. 118-33.

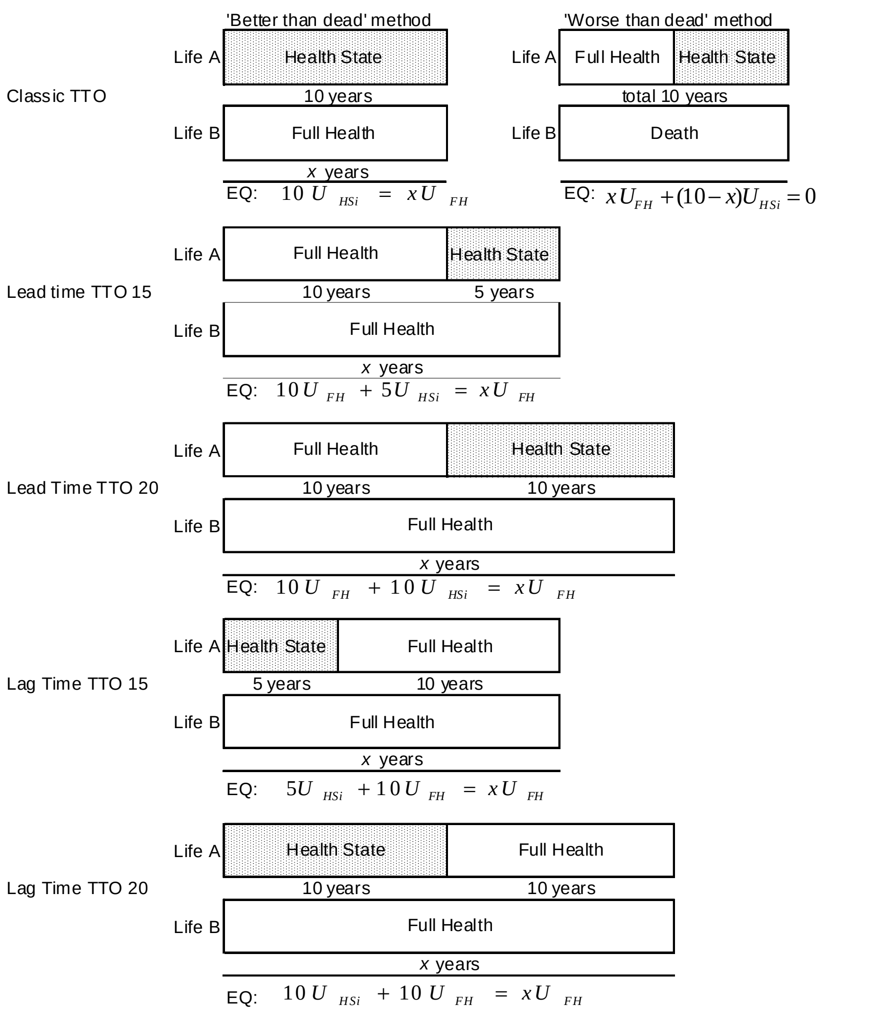
van Hout, B., Janssen, M.F., Feng, Y., Kohlman, T., Busschbach, J., Golicki, D., Lloyd, A., Scalone, L., Kind, P. & Pickard, A.S. In press, "Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets.*Value in health,* In Press.

Versteegh, M.M., Leunis, A., Uyl-de Groot, C.A. & Stolk, E.A. "Condition-Specific Preference-Based Measures: Benefit or Burden?", *Value in Health,* , 2012 Online early view.

Figure legend:

Table legend:

Figure 1: TTO specifications



U = utility of health state, x= years of full health at indifference
EQ: Utility equation.

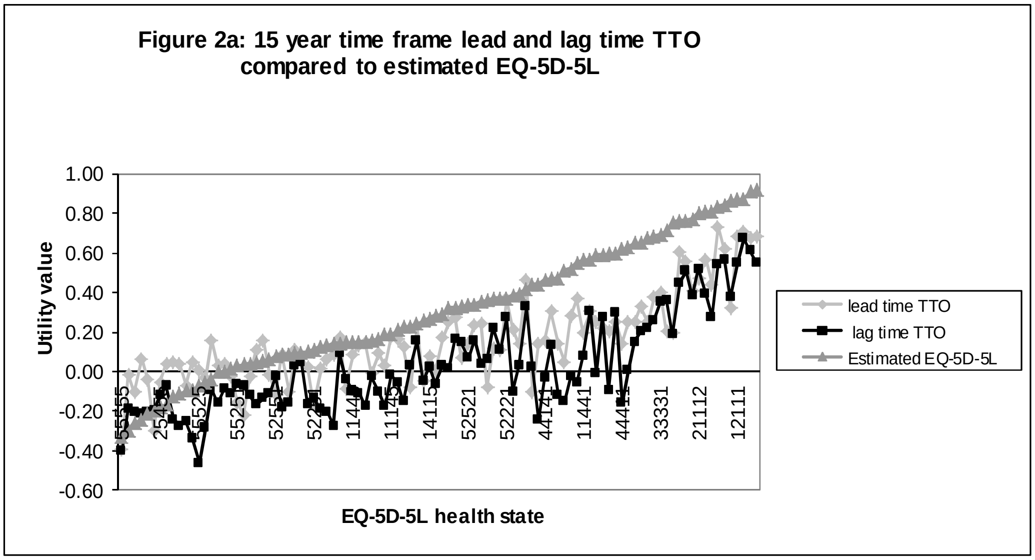


Figure 2a: 15 year time frame lead and lag time TTO
compared to estimated EQ-5D-5L
Utility value
1.00
0.80
0.60
0.40
0.20
0.00
-0.20
-0.40
-0.60
55555
25455
55525
55251
52551
52251
11445
11145
14115
52521
52221
44141
11441
44411
33331
21112
12111
EQ-5D-5L health state
lead time TTO
lag time TTO
Estimated EQ-5D-5L

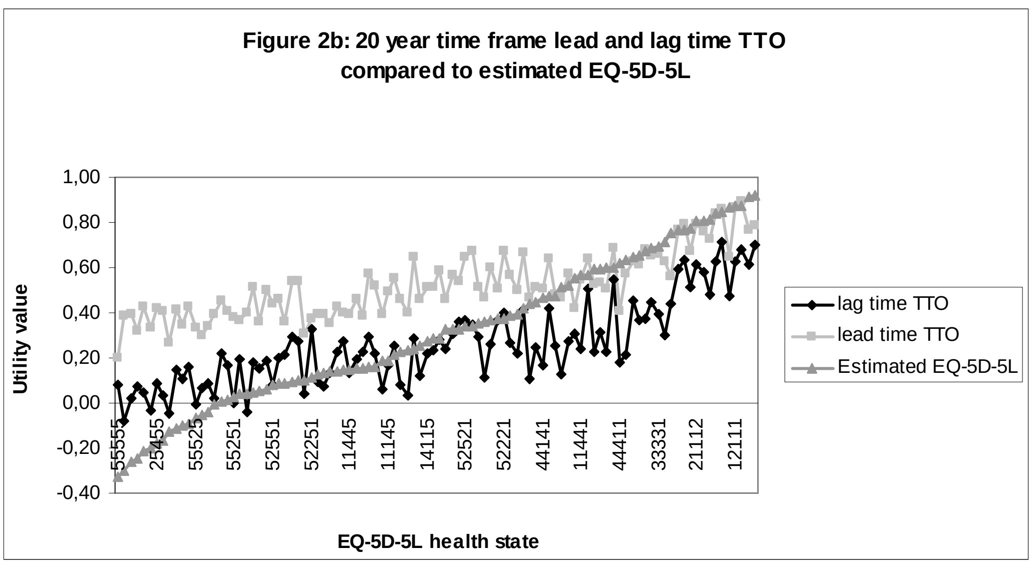


Figure 2b: 20 year time frame lead and lag time TTO
compared to estimated EQ-5D-5L
Utility value
1,00
0,80
0,60
0,40
0,20
0,00
-0,20
-0,40
55555
25455
55525
55251
52551
52251
11445
11145
14115
52521
52221
44141
11441
44411
33331
21112
12111
EQ-5D-5L health state
lag time TTO
lead time TTO
Estimated EQ-5D-5L

Figure 3: Standard deviations

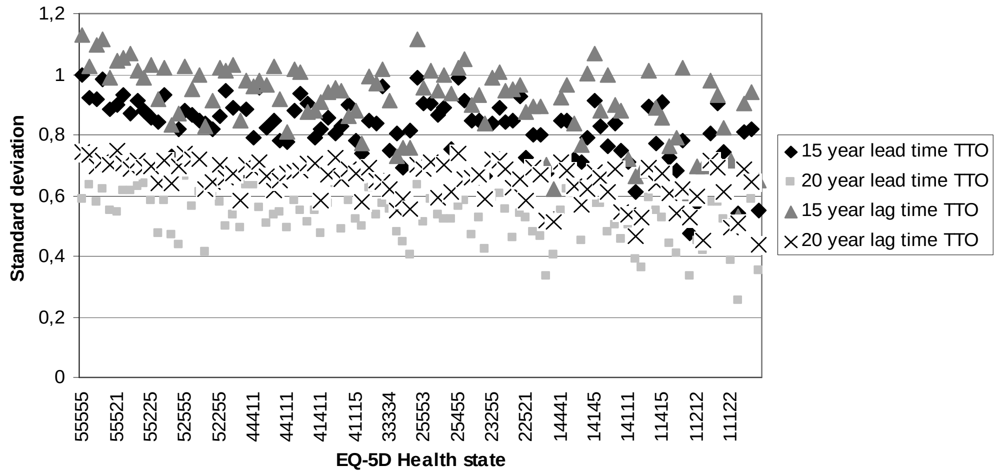


Table I: Checklist to compare TTO studies

	Comparator study	Online study	Different?
Type of TTO procedure	Lead time TTO	Lead time TTO / lag time TTO / classical TTO	No[1]
Total time frame	15 years	15/20 years for lead and lag time TTO, 10 for classical TTO	No[1]
Health state duration	5 years	5 / 10 years	No[1]
Lead time length	10 years	10 years	No
Lag time length	-	10 years	No[1]
Ratio of lead/lag time to health state duration	2:1	1:1 / 2:1	No[1]
Lowest possible value	-2	Lead time TTO: -2 / -1 lag time TTO: -2 / -1 Classic TTO: -19	No[1]
What was valued?	Health state description (EQ-5D-5L)	Health state description (EQ-5D-5L)	No
Who valued the states?	General population sample	General population sample	No
TTO procedure	Structured iteration	Structured iteration	No
Iteration first question	*t*=15 (utility = 1)	*t*=10/15/20 (utility = 1)	No
Iteration second question	*t*=10 (utility = 0)	*t*=0/10 (utility = 0)	No
Health state selection	D-Optimal design	D-Optimal design	No
Blocked design?	Yes	Yes	No
Number of health states	100	100	No
Per respondent	5	10	**Yes**
Number of attributes	5	5	No
Levels per attribute	5	5	No
Sample size	201	6222	**Yes**
Valuations per state	About 20	About 100	**Yes**
Time frame BTD	5 years	5/10 years	No[1]
Time frame WTD	10 (lead time)	10 (lead time)	No
Highest attainable value	1	1	No
Warm-up task / other tasks	Discrete choice experiment with other health states from same descriptive system	Discrete choice experiment and best worst scaling with other health states from same descriptive system	**Yes**
Visual presentation	Side by side presentation of alternatives	Side by side presentation of alternatives	No
	Striped colour bar as visual prop	Striped colour bar as visual prop	No
Standardized interview protocol?	Yes	Yes	No
Mode of administration	Self complete TTO preceded by interviewer instructions and interviewers on site	Online interviews with animated and textual assistance	**Yes**
Smallest tradable unit	3 Months	3 Months	No
Description of health state at which utility=1	Full health	Full health	No
Worst health state	Worst health state (55555)	Worst health state (55555)	No
WTD procedure	not applicable	not applicable	No

Transformation of WTD	not applicable	For classical TTO: Uhs/1-Uhs	No
Exclusion criteria	Excluded respondents who did not complete all tasks	Excluded respondents who did not complete all tasks	No

[1]Although the second on-line study applied more than just one TTO task, the lead time TTOs in the 15 year time frame were exactly the same.

Table II: Relative importance of different dimensions of health

	Classic TTO			15 year lead time TTO			20 year lead time TTO			15 year lag time TTO			20 year lag time TTO			Group TTO (15 year lead time)		
	Coef.	*p*	imp.	Coef.	*p*	imp.	Coef.	*p*	imp.	Coef.	*p*	imp.	Coef.	*p*	imp.	Coef.	*p*	imp.
Mobility	-0,026	***	3	-0,032	***	3	-0,026	***	3	-0,039	***	3	-0,036	***	3	-0,033	***	1
Self-care	-0,020	***	1	-0,027	***	1	-0,020	***	2	-0,033	***	1	-0,028	***	2	-0,040	***	3
Usual activities	-0,022	***	2	-0,028	***	2	-0,020	***	1	-0,038	***	2	-0,019	***	1	-0,038	***	2
Pain/Discomfort	-0,040	***	4	-0,057	***	5	-0,031	***	4	-0,060	***	5	-0,043	***	4	-0,087	***	5
Anxiety/Depression	-0,043	***	5	-0,053	***	4	-0,040	***	5	-0,058	***	4	-0,045	***	5	-0,073	***	4
Constant	0,731	***		0,740	***		0,915	***		0,692	***		0,751	***		0,872	***	
Adjusted R-square	0,12			0,13			0,10			0,12			0,13			0,24		

*** $p<0.01$

imp. = relative importance

Table III: Response characteristics

	%Utility = 1	%Utility = 0	%Utility < 0	%Utility = lowest value	% No differentiation between 10 health states	% Respondents using 4 or less iterations
Classic TTO	29,8	21,8	23,9	3,8	11,1	64,6
15 year lead time TTO	31,4	22,2	25,7	2,3	11,5	65,5
20 year lead time TTO	39,7	13,2	12,7	2,1	13,4	63,8
15 year lag time TTO	33,5	17,3	35,7	3,5	10,6	65,5
20 year lag time TTO	32,8	18,5	29,2	4,2	10,8	64,8
Group TTO	9,7	19,1	29,9	1,6	4,7	43,0

Table IV: Feasibility questionnaire

	S1			S2			S3			S4		
	Mean	[95% conf.int.]		Mean	[95% conf.int.]		Mean	[95% conf.int.]		Mean	[95% conf.int.]	
		lower	upper		lower	upper		lower	upper		lower	upper
Classic TTO	2,33	2,30	2,35	2,34	2,32	2,37	2,38	2,35	2,40	2,80	2,77	2,82
15 year lead TTO	2,32	2,29	2,35	2,29	2,27	2,32	2,37	2,35	2,39	2,73	2,71	2,76
20 year lead time TTO	2,67	2,64	2,70	2,69	2,67	2,72	2,38	2,36	2,41	2,97	2,94	3,00
15 year lag time TTO	2,40	2,37	2,43	2,39	2,37	2,42	2,38	2,36	2,41	2,81	2,78	2,83
20 year lag time TTO	2,39	2,37	2,42	2,36	2,34	2,39	2,36	2,33	2,38	2,86	2,83	2,88
Group TTO	1,51	1,47	1,54	1,66	1,62	1,70	2,27	2,22	2,32	2,21	2,16	2,26

Statement 1: The instructions that were given made it clear what I needed to do.

Statement 2: It was easy to understand the questions I was asked.

Statement 3: I found it difficult to decide on the exact point where life A and B were about the same.

Statement 4: I found it easy to tell the difference between the health states I was asked to think about.

Statement answer categories: 1 (totally agree) to 5 (totally disagree)