Universidade Vigo

Departamento de Economía Aplicada
http://webs.uvigo.es/x06

Working Paper: 11/01. November 2011

The Lead Time Trade-Off: The case of health states better than death

José Luis Pinto-Prades
Eva Rodríguez-Míguez

Coordinator: Eva Rodríguez Míguez
emiguez@uvigo.es

THE LEAD TIME TRADE-OFF:

THE CASE OF HEALTH STATES BETTER THAN DEATH

José Luis Pinto-Prades
Universidad Pablo de Olavide
Sevilla
Spain

Eva Rodríguez-Míguez*
Universidad de Vigo
Pontevedra
Spain

ABSTRACT

The Lead Time Trade-Off (L-TTO) is a variant of the TTO method which attempts to overcome some of the problems of the most widely used method for health states worse than death (SWD). Theoretically, the new method reduces the problems detected when researchers have elicited preferences for SWD. However, several questions remain to be clarified. One of them is the influence of this new method for states better than death (SBD). This paper attempts to shed some light on this issue using a split-sample design (n=500). One subsample (n=188) was interviewed using L-TTO and the rest using the traditional TTO (T-TTO). The results show that the L-TTO produces utilities that are consistently higher than the T-TTO for SBD. Furthermore, the greater the severity, the greater is the difference between both methods. Another finding is that the L-TTO seems to produce a lower number of SWD. This effect seems to be concentrated in the most severe health states. This implies a violation of additive separability, one of the cornerstones of the QALY model. The data show that the L-TTO may be different from the T-TTO in more respects than those that were originally intended.
Keywords: Lead TTO, states better than death, discounting.

* **Corresponding author**: Departamento de Economía Aplicada. Universidade de Vigo. As Lagoas-Marcosende s/n. 36212 Vigo. Spain. E-mail: emiguez@uvigo.es. Phone: +34-986813529.

Funding sources: Financial support from CIEF (Fundación Caixa Galicia), Consellería de Economía e Industria - Xunta de Galicia (10SEC300038PR), Ministerio de Ciencia e Innovación (ECO2008-05771 and ECO2010-2241-C02) and Proyecto de Excelencia de la Junta de Andalucía (PO7-SEJ-02936).

1. INTRODUCTION

One of the main components of any Economic Evaluation of Health Care Programmes is the utility of health states. They are very often estimated using the Time Trade-Off (TTO) (Torrance, Thomas & Sackett, 1972). Usually, utilities are elicited for chronic health states and they are applied to all sorts of health problems (chronic or temporary) and durations (long or short). Torrance (1986) presented two versions of TTO for chronic health states, one for states better than death (SBD) and one for states worse than death (SWD). The method Torrance (1986) proposed for SWD has been widely used (Tilling et al., 2010). A modified version of this method was used by the Euroqol Group in the UK Measurement and Valuation of Health (MVH) study (Dolan, 1997). However, it has been pointed out (Robinson & Spencer, 2006) that the framing used for SWD is very different from the framing used for SBD. In the case of SBD, subjects are asked to trade-off more years in bad health with fewer years in better (or full) health. In the case of SWD, they are asked to estimate the combination of years in full and bad health that is equivalent to death. Strictly speaking, if the assumptions of the QALY model held, and subjects had well-structured preferences for health problems, this change in framing should not be problematic. However, according to Robinson and Spencer (2006, p. 394) "there is a large body of evidence which shows that responses can be affected by simple variations in question wording – descriptive invariance - and the method used to elicit preferences – procedural invariance. Such evidence must call into question the validity of aggregating better than and worse than dead scores, generated by two different procedures". For these and other reasons it is perfectly reasonable to look at these two procedures as different "conceptually and operationally" (Devlin et al., 2011). Robinson and Spencer (2006) proposed a variant of the traditional TTO (henceforth T-TTO), namely, the Lead TTO (L-TTO). The L-TTO includes a certain number of years (L) in full health before the period in bad health. In L-TTO, utilities are estimated for SBD and SWD using the same procedure, that is, subjects have to indicate if they want to live longer with lower quality of life or vice versa.

However, the main reason to use L-TTO instead of T-TTO for SWD cannot just be that it avoids procedural invariance. If two procedures produce different results we cannot *solve* the problem by choosing one of them at random. The ultimate reason to choose L-TTO to elicit preferences for SWD has to be that utilities elicited with L-TTO are closer to what we can call the "true" utility. Tilling et al. (2010) and Devlin et al. (2011) provide some reasons that can be interpreted as a justification that utilities elicited with L-TTO are closer to "true" preferences or, equivalently, that T-TTO produces biased utilities for SWD. They argue that T-TTO "produces 'extreme' negative values" (Devlin et al, 2011, p. 349) for SWD. The fact that T-TTO produces 'extreme' values is not a problem per se if they reflect what people really think. We understand they are claiming that these "extreme negative values" do not reflect what people really think ("true" preferences). These values would be an artefact of the method. Also, they claim (Devlin et al, 2011, p. 359) that it is easier for people to " 'flip' from positive to negative values without the focusing effect created by the introduction of a separate valuation procedure". They seem to suggest that this "focusing effect" produce biased utilities. In summary, the argument seems to be that utilities provided by L-TTO are closer to the "true" value than those provided by T-TTO.

These arguments have a potential problem, namely, L-TTO and T-TTO could be producing different values for the same health state because the introduction of a lead period affects TTO in different ways than those initially envisaged. One possibility is that people may violate additive separability. If there are interactions between "disjoint time periods" (Wakker, 1996), adding a lead period can change the "true" utility of a health state. According to Devlin et al. (2011. p.359) "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". If this assumption does not hold, the "true" value of health states is bound to be different between T-TTO and L-TTO even if no biases are present. In this case, it cannot be said that utilities elicited with L-TTO ($U_{L\text{-}TTO}$) for SWD better represent preferences than utilities elicited with T-TTO ($U_{T\text{-}TTO}$) since preferences are not constant. Another possibility also mentioned by Devlin et al. (2011, p. 360) is that "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."

The objective of this paper is to find out to what extent the potential discrepancy between $U_{L\text{-}TTO}$ and $U_{T\text{-}TTO}$ can be attributed to violations of procedural invariance or to some other reason. This has been done by focusing on SBD. Since the procedure used to elicit utilities for SBD is the same under L-TTO and T-TTO, no violations of procedure invariance can explain a potential discrepancy between L-TTO and T-TTO for SBD.

2. COMPARING L-TTO AND T-TTO

In the T-TTO, at least for the chronic case of SBD, the utility of a health state is obtained after establishing indifference between two health profiles. Each profile is characterized by a combination of quality of life and time. That is, we get:

$$U(\boldsymbol{X},F;\ death)=U(\boldsymbol{T},B;\ death) \qquad [1]$$

Where, traditionally, F indicates full health, B is a SBD, $\boldsymbol{X}$ is the number of years in full health and $\boldsymbol{T}$ is life expectancy.

Usually $\boldsymbol{T}$ is fixed and $\boldsymbol{X}$ is adjusted ($\boldsymbol{X}<\boldsymbol{T}$) until indifference is reached. Under the usual scaling assumptions and applying the linear QALY model

$$U_{T\text{-}TTO}(B)= \boldsymbol{X}/\boldsymbol{T}. \qquad [2]$$

For SWD this method cannot be applied given that for these states there is no $\boldsymbol{X}>0$ which verifies [1]. The method developed by Torrance (1986) for SWD (denoted by W) estimates $U_{T\text{-}TTO}(W)$ from next indifference:

$$U(\boldsymbol{X},F;\ (\boldsymbol{T}-\boldsymbol{X}),W;\ death)=U(death) \qquad [3]$$

as follows

$$U_{T\text{-}TTO}(W)=-\boldsymbol{X}/(\boldsymbol{T}-\boldsymbol{X}) \qquad [4]$$

It is clear that [1] and [3] imply a very different task for the subject, so descriptive and procedural invariance can be easily violated. The L-TTO includes a certain number of years ($\boldsymbol{L}$) in full health before the period in bad health (H). That is, $U_{L\text{-}TTO}$ (H) is obtained by establishing indifference between the next two profiles:

$U(\boldsymbol{L},F;\boldsymbol{X},F;death) = U(\boldsymbol{L},F;\boldsymbol{T},H;death)$ [5]

As can be seen [1] and [5] are extremely similar. The only difference is that we add a common element ($\boldsymbol{L}$,F) to both sides of the equation. Since the QALY model assumes additive separability, adding this common element produces no difference in the utility of H, that is $U_{L\text{-}TTO}(H)$ should be equal to $U_{T\text{-}TTO}(H)$.

Let us define $\boldsymbol{Y}$ such as $\boldsymbol{Y}=\boldsymbol{L}+\boldsymbol{X}$, that is, the number of years in full health that are equivalent to $\boldsymbol{L}$ years in full health plus $\boldsymbol{T}$ years in bad health. Equality [5] can then be written as

$U(\boldsymbol{Y},F;death) = U(\boldsymbol{L},F;\boldsymbol{T},H;death)$ [6]

Under the linear QALY model and under the assumption of additive separability, $U_{L\text{-}TTO}$ (H) is estimated as

$U_{L\text{-}TTO}(H) = (\boldsymbol{Y}\text{-}\boldsymbol{L})/\boldsymbol{T}$ [7]

Observe than the framing in [6] can generate both positive and negative values for $U_{L\text{-}TTO}$ (H). This was not possible under [1].

The procedure to elicit utilities for SBD is basically the same under T-TTO and L-TTO ([7] and [2] are both $\boldsymbol{X}/\boldsymbol{T}$). In both cases, the subject is asked to seek indifference between two health profiles, one of them with lower life expectancy and the other with lower quality of life. The only difference between both framings is the common lead period in full health that is added to both profiles. This leads to the main hypothesis to be tested in this paper, namely, that the only difference between T-TTO and L-TTO is the different procedure they use. In order to test this hypothesis, two predictions are made:

a. Utilities for SBD are not systematically different between T-TTO and L-TTO.
b. The probability that a health state is considered better or worse than death does not change systematically between T-TTO and L-TTO.

If these hypotheses do not hold we will have evidence that the only difference between T-TTO and L-TTO is not the different procedure they use. If this were the case it would not be so straightforward to accept that utilities elicited with L-TTO for SWD are just the same utilities that the T-TTO tries to elicit, but estimated with less bias. It could imply that the lead period introduces other elements that modify the utilities estimated.

Two potential candidates that can introduce a discrepancy between $U_{T\text{-}TTO}$ and $U_{L\text{-}TTO}$ (apart from procedure) are violations of additive separability and discounting. If there are interactions between disjoint time periods, the introduction of the lead time in full health can modify how people perceive the severity of a health state, creating a discrepancy between $U_{T\text{-}TTO}$ and $U_{L\text{-}TTO}$. However, there is no theory that predicts the direction of this potential discrepancy. It is not the same with discounting. Under the constant discounting model (widely used in Economic Evaluation), the introduction of a lead period cannot explain any discrepancy between L-TTO and T-TTO for SBD since it assumes stationarity (see appendix). However, the literature has shown that this assumption is frequently violated and that temporal preferences can be better described assuming decreasing time aversion –DTA- (van der Pol and Cairns, 2002). It is shown in the appendix that these preferences could produce a discrepancy between $U_{T\text{-}TTO}$ and $U_{L\text{-}TTO}$ if responses to T-TTO and L-TTO are analysed using (wrongly) the linear QALY model. More specifically, it is shown that if U(H) is constant across contexts [$U_{T\text{-}TTO}(H)=U_{L\text{-}TTO}(H)$], temporal preferences are characterized by DTA, and however the responses to T-TTO and L-TTO are analysed, using the linear QALY model we would get $U_{T\text{-}TTO}(H)<U_{L\text{-}TTO}(H)$.It will be determined if these types of temporal preferences can explain the results.

In order to test our main hypothesis we conducted a survey in Galicia (North West of Spain) with 500 members of the general population. We now describe the survey and the statistical techniques. We then present the results and discussion closes the paper.

3. METHODS

3.1 Selection of health states

The survey used in this study was funded mainly in order to estimate utilities for health states associated with different levels of dependency generated by health problems. The descriptive system is showed in Table IIt gives rise to 1728 possible health states. The OPTEX Procedure from SAS Software (version 9.1) was used to generate a set of 24 health states divided into four blocks of six (Table II). Each participant in the survey valued only one of the four blocks (6 health states). Blocks were randomly allocated among subjects. We also randomized order of presentation of health states. Each participant only used T-TTO or L-TTO, that is, we used a between-sample design.

Table I: Dependency states. Brief description of attributes and levels

Eat	1. Does not need assistance to eat or drink. 2. Needs partial aid to eat or drink (cutting, serving, etc.). 3. Needs to be given food and drink.
Incontinence	1. Does not have incontinence or does not need help. 2. Has urinary incontinence (not faecal) and needs help for hygiene. 3. Has both urinary and faecal incontinence and needs help for hygiene.
Personal care	1. Does not need help for personal care: bathing, dressing, etc. 2. Needs help only to bath but not for the rest of his/her personal care. 3. Needs help for most personal care activities. 4. Is incapable of carrying out personal care. Needs someone to substitute him/her in this activity.
Mobility	1. Moves independently. 2. Does not need help to move within the home but does out of home. 3. Needs help to move both in and out of home. 4. Is incapable of changing position. Bed-ridden or chair-ridden.
Housework	1. Does not need help to carry out housework (cleaning, food, etc.). 2. Needs daily help for housework. 3. Is incapable of carrying out most tasks at home.
Mental problems	1. Does not have mental impairment. Is not mentally impaired. 2. Needs assistance to manage money, medication or to take some common everyday decisions. Collaborative attitude with the care-taker. 3. Incapable of taking basic decisions. Cannot live alone. Does not collaborate but does not offer resistance. 4. Incapable of taking basic decisions. Does not collaborate and usually offers resistance to help.

Table II: Dependency states valued by block*

Block 1	211121 133334 122222 214232 313331 323433	Block 3	111112 113233 213322 222131 234431 334234
Block 2	111221 112132 112211 223234 234333 333122	Block 4	123121 212223 233432 314434 324332 333231

* The number indicates the level of each attribute following the order of Table I.

3.2 Selection of respondents

Subjects were selected using a four-stage cluster stratified random sampling with final adjustment to quotes by sex and age. The reference population was between 18 and 65 years old. We did not include older people because the life expectancy we used in the L-TTO (20 years) clearly exceeded their own. A total of 500 interviews were conducted: 312 participants responded to the T-TTO protocol and 188 participants to the L-TTO protocol. Interviews were conducted face-to-face by six trained interviewers.

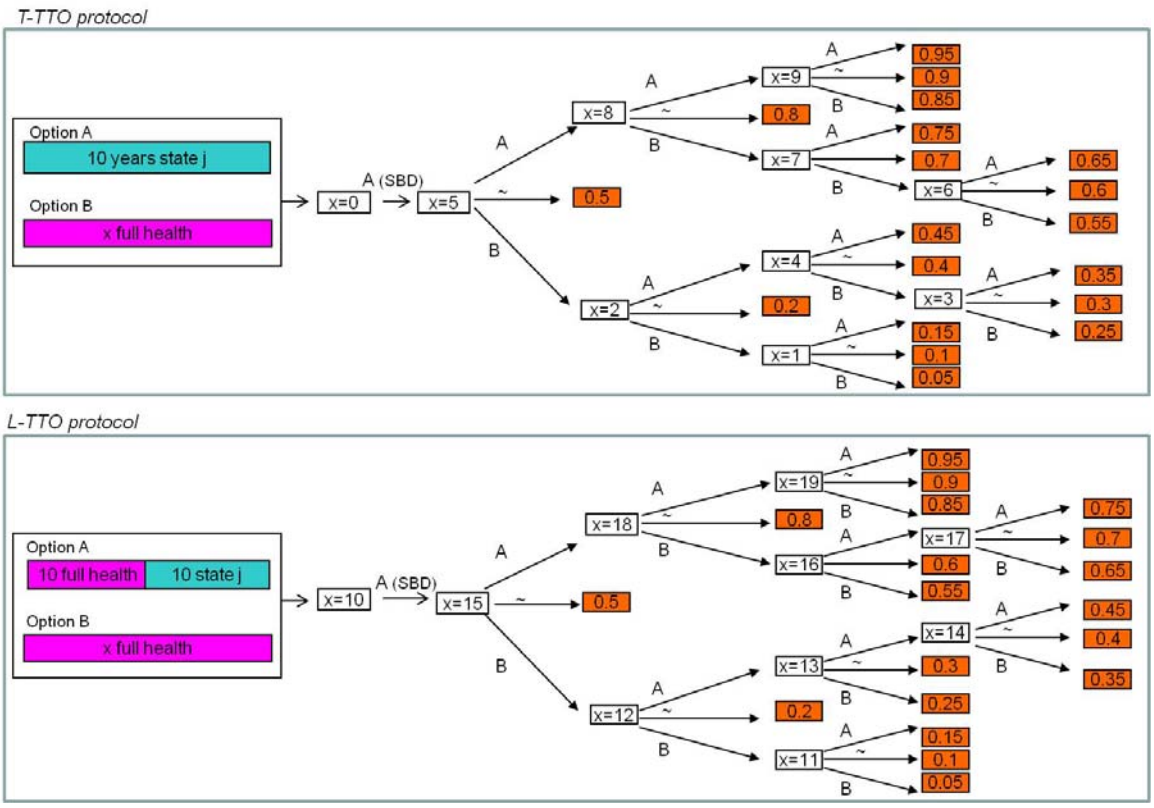
3.3 The questionnaires

We use two types of questionnaires, one for the T-TTO procedure developed by Torrance (1986) and another for the L-TTO procedure proposed by Robinson and Spencer (2006). Both types of questionnaire began by giving the motives for the study and an explanation of the health states (dimensions and levels) used in the questionnaire. Next, the subjects had to make a valuation of the six health states. We also collected the socio-demographic characteristics of the participants: age, gender, family income, education, labour status, living arrangements, size of municipality, own health (measured by Euroqol EQ-5D), whether they knew a dependent relative and also whether the relative lived with them.

3.4 Valuation procedure

The procedure is illustrated in Figure 1. Subjects had to choose between two options (A and B) with different health profiles. Visual aids were used to help the subject understand these questions. The first question aimed to identify if health states were considered better or worse than death. In order to find this, in the case of the T-TTO, the first question involved choosing between death in a few weeks and 10 years in a certain health state. In the case of the L-TTO the first question involved choosing between (10 years full health; death) and (10 years in full health; 10 years in bad health; death).

Figure 1: Protocol used in the questionnaire: T-TTO version and L-TTO version



Depending on the answer to the first question, the respondent followed a different path using a choice-bracketing procedure (series of ping-pong questions) as shown in Figure 1. In order to clarify the procedure, two examples are presented, one for the T-TTO and one for the L-TTO:

- T-TTO: Assume that somebody preferred (10 years, H; death) to death, then they would be asked to choose between (10 years, H; death) and (5 years, FH; death). The number of years in full health was moved up and down until an indifference interval (or value) was reached. When indifference was not directly obtained (most of the cases), we assigned the middle value of the indifference interval. For example, if (8 years, FH)$\succ$ (10 years, H) and (7 years, FH) $\prec$(10 years, H) we assumed that (7.5 years, FH)~(10 years, H) and, applying equation [2], U(H)=0.75. Figure 1 shows (in the shaded areas) the values assigned to the health states depending on the path followed by subjects.

- L-TTO: Assume that somebody preferred (10 years, FH; 10 years, H) to (10 years, FH; death), then they would be asked to choose between (15 years, FH; death) and (10 years, FH; 10 years, H; death). Using the choice-bracketing technique shown for the T-TTO, we obtained an indifference (or value) interval. If (10 years, FH; 10 years, H; death) ~(12 years, FH ; death) applying equation [7] we would have that U(H)=0.2. Figure 1 shows (in the shaded areas) the values assigned to the health states depending on the path followed by subjects.

3.5 Analysis

Consistency of the responses

The violations of dominance were analysed in order to test the consistency of responses. It was considered that a health state dominates another if it is at least better in one dimension and it is not worse in any of the other dimensions. As can be checked in Table II, there are several dominance situations (6 in the blocks 1-3 and 4 in the block 4). For instance, in block 1 the health state 313331 dominates 323433. The number of participants was identified who did not verify dominance at least once in both protocols.

Hypotheses testing

The two hypotheses were tested as follows:

1. To test if the utilities for SBD depend on the protocol, we formulated the following model:

$$U_{ij} = \alpha + \sum\beta_j s_j + \delta' x_i + \gamma Lead + \varepsilon_{ij}, \quad [8]$$

where U_{ij} is the utility assigned by respondent i to the health state j (j = 1, 2, … , 24) if U_{ij}>0 (obtained applying the equation [2] for the T-TTO sample and the equation [7] for the L-TTO sample); s_j is a dummy variable indicating the state valued (e.g., $s_j = 1$ if j=1 and $s_j = 0$ if $j\neq1$); x_i is a vector of personal characteristics of the participants; *Lead* is a dummy variable indicating if the participant used the L-TTO protocol (*Lead*=1) or the T-TTO protocol (*Lead*=0); εij is an error term and α, β_j, δ' and γ are the parameters to be estimated. This model was estimated using the random effects regression model because it takes into account that if the same individual values several health states**,** then the observations provided by that same participant cannot be considered independent. This model considers that $\varepsilon_{ij} = u_j + e_{ij}$ where u_j is the individual specific error term and e_{ij} is the traditional error term associated to each observation. We test if γ is statistically different from zero to test if L-TTO and T-TTO produce systematically different results.

2. To test if the probability that a health state is considered better or worse than death changes or not systematically between T-TTO and L-TTO, we estimate a random effect logit model. The independent variables are the same as in [8]. The dependent variable is binary, taking a value of 1 if the respondent considered this state worse than death and 0 otherwise. We estimated a random-effects logit model in order to capture unobserved factors specific to each respondent. Finally we tested if γ is statistically different from zero to analyse if the format used affects the probability that a state will be considered worse or better than death.

4. RESULTS

Table III shows the characteristics of respondents in both L-TTO and T-TTO samples. Both samples present similar socioeconomic characteristics and seem to have similar preferences regarding the importance of attributes. There are two fairly homogeneous samples. This suggests that the results regarding comparison between both methods can be robust. In any case, the analysis conducted was controlled for potential differences between samples.

Table III: Characteristics of respondents by type of questionnaire (%)

		L-TTO N=188	T-TTO N=312
Sex	Female	55.8	47.4
Age	Mean	40.9	41.5
Education	Primary studies or less	35.1	37.5
	Secondary	37.2	39.4
	University	27.7	23.1
Habitat	Rural	34.6	31.4
	Intermediate	29.3	31.1
	Urban	36.2	37.5
Living alone		9.7	13.5
Good health	(EQ-5D=11111)	68.6	76.3
Know	Any close dependent	31.4	53.2
	Close dep. (not live together)	59.0	40.1
	Close dep. (live together)	9.6	6.7
Labour status	Employed	58.0	59.6
	Pensioner/retired	6.4	10.9
	Unemployed	23.4	16.0
	Student	6.4	5.1
	Domestic tasks	5.9	8.3
Home income (€ monthly)	<=500	6.1	5.9
	500-1000	23.9	13.2
	1000-1500	25.0	30.5
	1500-2000	16.7	25.7
	2000-3000	20.0	16.9
	>3000	8.4	7.7
Duration of interview (minutes)		22.5	23,2
Participants who placed it in first place	Eat	4.8	8.0
	Incontinence	5.9	7.1
	Personal care	4.3	6.7
	Mobility	7.5	8.7
	Housework	0.0	0.3
	Mental	77.7	69.2
Participants who placed it in second place	Eat	16.49	16.99
	Incontinence	45.21	30.13
	Personal care	10.11	19.55
	Mobility	15.43	20.19
	Housework	2.13	2.88
	Mental	10.64	10.26

Firstly the consistency of the responses was tested. This was done by analysing dominance. Since there are 22 pairwise combinations of states where one state dominated the other (it was not worse in any dimension and it was better in some other(s) dimension (s)) we tested if the parameters were significantly different from each other in these pairwise comparisons. The hypothesis of equality of parameters is rejected at the 5% level in 18 pairs of states and always in the right direction, that is, the parameter associated with the dominant health state was always higher.

The main results regarding the two hypotheses can be seen in Table IV. In both cases two regressions were run, excluding (models 1 and 4) or including (models 2 and 5) personal characteristics[1]. Hypothesis 1 is clearly rejected in both models (see model 1 and 2). The coefficient of the *Lead* variable is positive and significantly different from zero. It is also quite high since the L-TTO adds about 0.2 points to the average utility of health states, with regard to the T-TTO method. Since we are dealing with states that move between 0 and 1 this is a very important effect.

As mentioned in section 2 (and shown in the appendix) our results could be explained by DTA. If we assume (erroneously) a linear QALY model and temporal preferences are characterized by DTA we would get $U_{T\text{-}TTO}<U_{L\text{-}TTO}$ even if they are really the same. Therefore we should test if these differences are eliminated by applying a DTA model. We used Harvey (1986) and Mazur (1987) models since they have been used in the health economics literature (van der Pol & Cairns, 2002). In Harvey (1986) $\delta_t=1/(1+t)^h$, and in Mazur (1987) $\delta_t=1/(1+gt)$. Utilities were estimated using these two models with the parameters estimated by van der Paul and Cairns (2002) in a health context (h=0.25 and g=0.15). The coefficient of the *Lead* variable in the model 1 was 0.15, which is smaller than that previously estimated, but it was still statistically significant at the 1% level and quite large. A sensitivity analysis was also conducted using a wide of range of values for parameters *h* and *g,* and the coefficient of the *Lead* variable was always statistically different from zero. In summary, discounting does not seem to be the fundamental explanation of the difference between T-TTO and L-TTO. Violations of additive separability are left as the main candidate.

[1] Income was excluded from the analysis because 9.6% of subjects did not respond to this question. The models were estimated including this variable and the results did not change.

Table IV: Results of the estimation

	Hypothesis 1 Random regression model			Hypothesis 2 Random Logit model		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Constant	0,637**	0,646**	0,651**	-4,596**	-6,544**	-4,786**
States [Ref: 211121]						
133334	-0,249**	-0,247**	-0,315**	6,946**	6,935**	7,453**
122222	-0,196**	-0,195**	-0,195**	2,741**	2,718**	2,755**
214232	-0,186**	-0,184**	-0,215**	3,650**	3,619**	3,768**
313331	-0,130**	-0,128**	-0,161**	3,032**	3,006**	3,149**
323433	-0,252**	-0,251**	-0,324**	6,682**	6,665**	7,183**
111221	0,048	0,053	0,049	0,218	0,155	0,173
112132	-0,096**	-0,091**	-0,095**	2,013**	1,965**	1,975**
112211	0,015	0,020	0,017	0,218	0,155	0,173
223234	-0,233**	-0,229**	-0,300**	6,769**	6,709**	7,299**
234333	-0,299**	-0,294**	-0,368**	6,330**	6,274**	6,853**
333122	-0,200**	-0,195**	-0,231**	5,767**	5,716**	5,855**
111112	-0,041	-0,041	-0,037	1,052	0,916	0,981
113233	-0,171**	-0,171**	-0,202**	5,046**	4,927**	5,106**
213322	-0,206**	-0,206**	-0,237**	4,780**	4,661**	4,840**
222131	-0,146**	-0,146**	-0,144**	2,778**	2,655**	2,708**
234431	-0,248**	-0,250**	-0,280**	6,498**	6,390**	6,558**
334234	-0,271**	-0,272**	-0,330**	7,500**	7,407**	8,071**
123121	-0,098**	-0,096**	-0,096**	2,143**	2,200**	2,095**
212223	-0,142**	-0,138**	-0,176**	5,348**	5,427**	5,431**
233432	-0,185**	-0,182**	-0,247**	6,713**	6,790**	7,258**
314434	-0,191**	-0,188**	-0,258**	7,877**	7,954**	8,436**
324332	-0,188**	-0,184**	-0,242**	6,618**	6,695**	7,162**
333231	-0,151**	-0,148**	-0,187**	5,194**	5,274**	5,277**
Lead [Ref: T-TTO]	0.179**	0.204**		-0.400	-0.498	
Lead2 [Ref: T-TTO]						
Group 1 (less severe)			0.136**			0.230
Group 2 (intermediate)			0.227**			-0.084
Group 3 (more severe)			0.281**			-1.298**

Table IV: Results of the estimation (cont.)

Sex [Ref: female]		-0.041		-0.059
Age		0.001		0.013
Education [Ref: primary]				
Secondary		0.006		-0.051
University		-0.065*		0.257
Habitat [Ref: rural]				
Intermediate		-0.050		1.311**
Urban		0.025		1.298**
Living alone [Ref:No]		0.013		0.065
Know [Ref: Any close ...]				
Not live together.		-0.079**		0.350
Live together.		-0.140**		0.068
Good health [Ref: EQ-5D≠11111]		0.036		0.269
Labour status [Ref:employ.]				
Pensioner/retired		-0.041		-0.435
Unemployed		0.026		0.676*
Student		-0.055		0.770
Domestic tasks		0.066		0.071
Respondents	456	456	500	500
Observations	1557	1557	3000	3000

**Significant at the 5% level; *significant at the 10% level

If the explanation of the result is that there are interactions between disjoint time periods, it seems logical to think that this effect may depend on the severity of the health state. In order to test this hypothesis, health states were divided into three groups according to severity. Each group had 8 health states: the less severe states were in group 1 and the most severe in group 3. The severity of a state was approximated according to the proportion of participants that considered the state as better or worse than death in the T-TTO. Although it seems natural to identify the severity according to the utility of the health state, this had the problem that for some health states the number of observations was small, since most people considered the state as worse than death. Model 3 shows that the difference between L-TTO and T-TTO increases with severity. There are significant differences (Wald test) between the parameters of groups 1 and 2 and 1 and 3 at the 5% level and between groups 2 and 3 at the 10% level.

Hypothesis 2 cannot be rejected in either of the two models for the whole sample. The *Lead* variable is not significantly different from zero in model 4 and 5, indicating that the probability that a state is considered worse or better than death is not different between the T-TTO and the L-TTO. However, model 6 shows that the probability that a health state is considered worse than death is lower with L-TTO for the most severe health states. Adding a lead time seems to have a special influence on the utility of the most severe health states. This result also seems to support the conclusion that the disparity between T-TTO and L-TTO can be produced by a violation of additive separability since discounting cannot change the consideration of a health state as better or worse than death.

Other auxiliary regressions were also conducted (results not shown) in order to test the stability of the results. Including socio-demographic variables, including and excluding missing values related with income, and excluding participants who failed dominance at least once did not change the main results.

5. DISCUSSION

The main results of this study are that: a) L-TTO seems to produce higher utilities than T-TTO for SBD, b) this effect seems to increase with severity, c) the probability that a heath state is considered as better or worse than death is only different between both methods for the most severe states. We conclude that the L-TTO and the T-TTO produce different utilities for SBD.

These results seem to reflect a violation of the principle of additive separability, that is, people perceive health states differently if a lead period in full health is added upfront. There is nothing wrong in violating this assumption. It is a convenient assumption (it makes the QALY model more tractable) but it is not a normative assumption. As Wakker (1996) has pointed out, QALY assumptions can only be expected to hold approximately and “whether the greater tractability of analysis outweighs the loss of empirical realism is a question that cannot be answered in a universal manner; the answer depends on context and application” (p. 209). In our context, assuming additive separability seems to have a high cost in terms of empirical realism.

One explanation of this result is that the introduction of a lead period in full health allows people to prepare for the bad years that will come. While in T-TTO the bad years are a surprise (they start immediately), in L-TTO people have time to make adjustments. If 10 healthy years are added upfront, a person can make preparations during those 10 years for what is to come after, reducing burdens and hence diminishing differences. The important jobs to be done can now be taken care of before the trouble starts.. This is a kind of interaction between disjoint time periods that is very natural. If this is true, the consequence is that the “true” utility that T-TTO and L-TTO elicit is intrinsically different. For this reason, if both methods produce different results for SWD we cannot conclude that L-TTO produce utilities that are closer to “true” values since it seems that “true” values are different. They are context-dependent. For example, in those contexts where illnesses are diagnosed in advance and symptoms do not show up immediately (e.g. Parkinson), utilities elicited with L-TTO can be closer to “true” preferences.

As far as we know, there are at least two more studies that have presented some evidence on the issues that we have explored in this paper. Devlin et al. (2011) interviewed a group of 109 subjects from the general population using the L-TTO. They compared their results with those of the MVH study and their conclusions were similar to ours. That is, they found that L-TTO and T-TTO produced comparable proportions of respondents considering a health state better or worse than death in 7 out 10 states. In those states where there were significant differences, the two more severe states (EQ-5D 13332 and EQ-5D 23232) L-TTO produced a lower proportion of respondents considering this state worse than death, and the opposite occurs for the less severe state (EQ-5D 11112). Also, they found that in four of the 10 health states analysed, the L-TTO produced higher utilities than T-TTO for SBD. Attema et al. (2011) also compared T-TTO and L-TTO. They valued only six health states but they used several lead times. They show that the utility of health states depend on the lead time. Utilities were lower for a shorter lead time (5 years) than for a longer (10 and 17 years) in the case of SBD. This represents another violation of additive independence. In order to compare results, we will only focus on their results corresponding to the lead time that we used (10 years). Unfortunately, they only used 3 health states with this lead time and their results are inconclusive. For the mildest health state (EQ-5D 11121) the mean (and median) utility of T-TTO was higher than the utility of L-TTO and for the intermediate health state (EQ-5D 11113) and the worse health state (EQ-5D 23232) L-TTO produced a higher value than T-TTO (the medians were the same in the intermediate state and higher in the L-TTO for the worse state).

There is a message that seems to come from these studies that have used L-TTO, namely, utilities are not constant across contexts. The ultimate solution to this problem is to try to understand how context influence utilities and use those values that better represent preferences in each context. However, this can be (or not) considered unfeasible and unrealistic. In the meantime, utilities have to be used in decision making. A decision has to be taken to either continue with T-TTO or move to L-TTO. Robinson and Spencer (2006) and Devlin et al (2011) have presented some arguments in favour of the L-TTO. What our paper suggests is, that if L-TTO is going to substitute T-TTO this would not only affect SWD but also SBD and a whole new set of values will have to be produced. These utilities will not just be better estimations of the same "true" values formerly elicited with T-TTO but a whole set of "new true values".

Our paper also has limitations that should be overcome in future research. We have compared both methods in a between-sample design. While we think this is the best design in order to test if both methods produce different utilities, it is not the best method in order to understand why T-TTO and L-TTO produce different results. This can be better addressed in a within-sample design. Of course, a within-sample design may have confounding factors of its own that will have to be carefully controlled (e.g. order and learning effects) but it seems the right design in order to understand the reason for the discrepancy. Another limitation of our study is that subjects were not randomized between both methods. Administrative and organizational issues made randomization impossible. Since the socio-demographic characteristics of both samples were quite similar and since we used multivariate analysis in order to control for biases coming from non-randomization, we do not think there are obvious reasons to suspect that our results are biased.

Recently Devlin et al., (2011. p. 348) suggested that one topic for further research with L-TTO was "to better understand the implications for valuations of states better than dead". Our paper is an attempt to fill this gap. Further research should try to understand why adding a common outcome or why changing the size of this common outcome to health profiles seems to influence the utilities elicited so heavily.

APPENDIX

Effect of discounting in T-TTO and L-TTO for SBD.

Assume somebody is indifferent between X years in full health (F) and T years in health state H ($X<T$, H<F), under T-TTO. The utility of H would be estimated as

$$U_{T-TTO}(\mathbf{H}) = \frac{(\delta_1 + \cdots + \delta_X)}{(\delta_1 + \cdots + \delta_X + \cdots + \delta_T)} \quad 0 < \delta_t < 1 \qquad [1a]$$

where δ_t is the weight associated to one life year that occurs in the period t ($\delta_t=1$ for the linear QALY model).

In L-TTO we add a common delay L in full health to both profiles. Utility is estimated as

$$U_{L-TTO}(\mathbf{H}) = \frac{(\delta_{L+1} + \cdots + \delta_{L+X})}{(\delta_{L+1} + \cdots + \delta_{L+X} + \cdots + \delta_{L+T})} \qquad [2a]$$

1. Under the constant discounting model the *relative* benefit of receiving one outcome sooner (t) or later (t') only depends on the absolute distance between t and t'. That is, if we delay t and t' by a common period L, we have that

   $$\frac{\delta_t}{\delta_{t'}} = \frac{\delta_{L+t}}{\delta_{L+t'}}$$

   Given that under constant discounting $\delta_{L+t}=\delta_L\times\delta_t$, then [1a]=[1b]. That is $U_{T\text{-}TTO}(H)=U_{L\text{-}TTO}(H)$

2. Under a decreasing discount the effect of a delay L (δ_L) is not constant but it increases with the moment in time it is produced. That is, if $t'>t$ we have that

   $$\frac{\delta_{t'}}{\delta_t} < \frac{\delta_{L+t'}}{\delta_{L+t}}$$

   Therefore

$$\sum_{t=1}^{X}\frac{1}{\delta_t}\left[\sum_{t'=1}^{T}\delta_{t'}\right] < \sum_{t=1}^{X}\frac{1}{\delta_{L+t}}\left[\sum_{t'=1}^{T}\delta_{L+t'}\right] \;\rightarrow\; \frac{\sum_{t=1}^{X}\delta_t}{\sum_{t'=1}^{T}\delta_{t'}} > \frac{\sum_{t=1}^{X}\delta_{L+t}}{\sum_{t'=1}^{T}\delta_{L+t'}}$$

That is $U_{T\text{-}TTO}(H)>U_{L\text{-}TTO}(H)$. Therefore if a subject gives up the same number of life years in full health in T-TTO as in L-TTO and they have temporal preferences characterized by DTA they would be implying that $U_{T\text{-}TTO}(H)>U_{L\text{-}TTO}(H)$.

Assume that somebody has preferences such that U(H) is constant across contexts, that is, $U_{T\text{-}TTO}(H)=U_{L\text{-}TTO}(H)$ and they have temporal preferences characterized by DTA. In this case they would give up less life years in L-TTO than in T-TTO. If this were the case and we analysed T-TTO and L-TTO responses using the linear QALY model then we would find that $U_{T\text{-}TTO}(H)<U_{L\text{-}TTO}(H)$.

REFERENCES

Attema AE, Versteegh MM, Oppe M, Brouwer W, Stolk EA. Lead time TTO: leading to better health state valuations?. Submitted for publication.

Devlin NJ, 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* **20**(3):348-61.

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

Harvey CM. 1986. Value functions for infinite period planning. *Management Science* **32**: 1123-1139.

Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. 1963. Studies of illness in the aged. The index of ADL: a standardized measure of biological and psychosocial function. *JAMA* **185**: 914-919.

Mahoney FI, Barthel DW. 1965. Functional evaluation: the Barthel index. *Maryland State Medical Journal* **14**: 61-5.

Mazur JE. 1987. An adjustment procedure for studying delayed reinforcement. In: Commons ML, Mazur JE, Nevins JA, Rachlin H. (Eds.). *Quantitative analysis of behaviour: the effect of delay and intervening events on reinforcement value*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Robinson A, Spencer A. 2006. Exploring challenges to TTO utilities: valuing states worse than dead. *Health Economics* **15**(4): 393-402.

Tilling C, Devlin N, Tsuchiya A, Buckingham K. 2010. Protocols for time trade off valuations of health states worse than dead: a literature review. *Medical Decision Making* **30** (5): 610-619.

Torrance GW. 1986. Measurement of health state utilities for economic appraisal. *Journal of Health Economics* **5**(1): 1-30.

Torrance GW, Thomas WH, Sackett DL. 1972. A utility maximization model for evaluation of health care programs. *Health Services Research* **7**(2): 118-33.

Van der Pol M, Cairns J. 2002. A comparison of the discounted utility model and hyperbolic discounting models in the case of social and private intertemporal preferences for health. *Journal of Economic Behavior & Organization* **49**(1): 79-96.

Wakker P. 1996. A criticism of healthy-years equivalents. *Medical Decision Making* **16**(3): 207-14.