Glasgow Caledonian University

University for the Common Good

The lead time tradeoff

Pinto Prades, Jose Luis; Rodríguez-Míguez, Eva

Published in:
Medical Decision Making

DOI:
10.1177/0272989X14541952

Publication date:
2015

Document Version
Peer reviewed version

The Lead Time Tradeoff: The Case of Health States Better Than Dead

José Luis Pinto-Prades and Eva Rodríguez-Míguez

Med Decis Making published online 9 July 2014
DOI: 10.1177/0272989X14541952

The Lead Time Tradeoff: The Case of Health States Better Than Dead

José Luis Pinto-Prades, PhD, Eva Rodríguez-Míguez, PhD

Background. *Lead time tradeoff (L-TTO) is a variant of the time tradeoff (TTO). L-TTO introduces a lead period in full health before illness onset, avoiding the need to use 2 different procedures for states better and worse than dead. To estimate utilities, additive separability is assumed. We tested to what extent violations of this assumption can bias utilities estimated with L-TTO.* **Methods.** *A sample of 500 members of the Spanish general population evaluated 24 health states, using face-to-face interviews. A total of 188 subjects were interviewed with L-TTO and the rest with TTO. Both samples evaluated the same set of 24 health states, divided into 4 groups with 6 health states per set. Each subject evaluated 1 of the sets. A random effects regression model was fitted to our data. Only health states better than dead were included in the regression since it is in this subset where additive separability can be tested clearly.* **Results.** *Utilities were higher in L-TTO in relation to TTO (on average L-TTO adds about 0.2 points to the utility of health states), suggesting that additive separability is violated. The difference between methods increased with the severity of the health state. Thus, L-TTO adds about 0.14 points to the average utility of the less severe states, 0.23 to the intermediate states, and 0.28 points to the more severe estates.* **Conclusions.** *L-TTO produced higher utilities than TTO. Health problems are perceived as less severe if a lead period in full health is added upfront, implying that there are interactions between disjointed time periods. The advantages of this method have to be compared with the cost of modeling the interaction between periods.* **Key words:** *cost utility analysis; lead time trade off; states better than dead; additive separability; random effects regression.* ***(Med Decis Making XXXX;XX:xx–xx)***

Time tradeoff (TTO)[1] is one of the main techniques used to estimate utilities for health states. Torrance[2] presented 2 versions of TTO for chronic health states, one for states better than dead (SBD) and another one for states worse than dead (SWD). For SBD, the utility of a health state H is estimated from an indifference between 2 health profiles such as X years in full health, (X,F), and T years in H, (T,H), both followed by death. From this, U(H) = X/T. The main assumption that we need to estimate U(H) is that the utility of duration is linear[3]; that is, U(T) = T and U(X) = X. This assumption ‘‘permits the use of the time trade-off technique for eliciting health states utilities.’’[3(p842)] It is not possible to use this framing if (X,F) is always preferred to (T,H) for any value of X (even if X = 0). In this case, the health state is considered worse than dead. The method developed by Torrance[2] for SWD involves establishing indifference between immediate death and a profile where some years (X) in full health are followed by some years (Z) in bad health (and death). At the point where individuals are indifferent between (X,F;Z,H) and immediate death, the utility of H is estimated as U(H) = $-X/Z$. In the case of SWD, we also need to assume that additive independence holds, since U(X,F;Z,H) is decomposed as U(X,F) + U(Z,H) to obtain U(H).

The procedures used for SBD and SWD are very different from a descriptive perspective. In the first

Received 31 May 2013 from Glasgow Caledonian University, Newton Mearns, UK (JLPP); and University of Vigo, Vigo, Spain (EMRM). Eva Rodriguez Miguez acknowledges financial support from the Spanish Ministerio de Ciencia e Innovación (grant number ECO2011-25661) and the Consellería de Economía e Industria Xunta de Galicia (grant number 10SEC300038PR). Jose Luis Pinto Prades acknowledges financial support from the Spanish Ministerio de Ciencia e Innovación (grant number ECO2010.22041.C02.01) and Junta de Andalucía (grant number P09.SEJ.4992). Revision accepted for publication 16 May 2014.

Supplementary material for this article is available on the *Medical Decision Making* Web site at http://mdm.sagepub.com/supplemental.

Address correspondence to José Luís Pinto-Prades, Glasgow Caledonian University, 5 Firwood Road, Newton Mearns, G77 5PX, UK; e-mail: joseluis.pintoprades@gmail.com.

DOI: 10.1177/0272989X14541952

case, subjects have to estimate the number of years in full health they are willing to *give up* in order to improve their quality of life so that both profiles produce the same utility. In the second case, they have to estimate the *increase* in the number of years in full health that compensate the negative value of life-years lived in H so that the addition of the 2 periods is equivalent to death. This has led some authors to ''call into question the validity of aggregating better than and worse than dead scores, generated by 2 different procedures.''[4(p394)] The evaluation of SWD with TTO has also been problematic for other reasons.[5] So the negative values can be extremely negative and the lowest possible score depends on the smallest unit of time used. To avoid the effect of these extreme values in the mean score, usually the negative values are transformed to a scale of –1 to 0. However, this transformation does not have a theoretical justification. In addition, the use of different procedures for SWD and SBD produces discontinuities around 0 (existence of a gap effect around dead). Finally, it has also been shown that SWD are more influenced by sequence effects than SBD.[6] To address some of these problems, Robinson and Spencer[4] proposed a variant of TTO, namely, lead TTO (L-TTO). This method is similar to TTO for SBD except that it includes a certain number of years (L) in full health that precede both (X,F) and (T,H). That is, individuals are asked to compare (L,F;X,F) and (L,F;T,H). In practice this method asks people to provide the number of years in full health, Y (being $Y = L + X$), such that U(Y,F) = U(L,F;T,H). In this case, the utility of H is estimated as U(H) = $(Y - L)/T$.

TTO for SBD and L-TTO use the same procedure to elicit utilities; namely, subjects have to give up years in full health in order to increase quality of life. However, contrary to TTO for SBD, L-TTO can generate positive and negative values for H since U(H) $\gtreqless$ 0 if $Y \gtreqless L$. It is not necessary to use a different procedure for SBD and SWD. This attractive feature of the method comes with some cost: It requires more assumptions than TTO for SBD in order to elicit utilities. In the case of TTO for SBD, the assumption that we need to estimate U(H) as X/T is that the linear quality-adjusted life-years (QALY) model holds. In the case of L-TTO, as Devlin and others[7] observed, the method ''relies on the assumption of additive separability''; that is, we need to assume U(L,F;T,H) = U(L,F) + U(T,H). The question is, how good is this assumption? How close is it to real preferences?

In the context of L-TTO, additive separability implies that subjects are indifferent between knowing and not knowing in advance (with certainty) that they will have a health problem in the future. However, this might not be the case. Some people might prefer not knowing in advance in order to avoid the anxiety generated by this knowledge, while others might prefer to know in order to make preparations for the period of bad health. In both cases, the utility of the whole profile is not the addition of 2 separate components since there are interactions between the 2 components.

The first objective of this paper is to test whether additive separability holds in the case of L-TTO. To conduct this test, we will compare utilities produced with TTO and L-TTO only for SBD. Our null hypothesis will be that utilities are not different. If they are, we suggest that the most plausible explanation is that additive separability is violated. A corollary of our null hypothesis is that the probability of considering a health state as worse than dead does not change systematically between TTO and L-TTO. This will also be tested.

The reason to focus on SBD to test additive separability is that there are fewer confounding factors in the case of SBD. A disparity between TTO and L-TTO for SWD could be attributed to other reasons apart from additive separability, namely, a different procedure or discounting. However, with SBD this is not the case. To see why, we could think that TTO for SBD is just a special case of L-TTO where $L = 0$. That is, L changes the moment in time where the tradeoff between years of life in full health and bad health is produced. Under exponential (constant) discounting, the relative value of 2 outcomes depends only on the distance between the events (years in full and bad health) and not on the moment in time where they are produced (L). It does not matter whether $L = 0$, or not as long as it is the same for the profile in bad health or full health. For instance, if a subject is indifferent between (5,F) and (10,H) and we assume, without loss of generality, a discount rate of 3%, then U(H) = 0.537. This utility will change with the discount rate but it will be the same in TTO or L-TTO for all values of L and all discount rates. This does not happen for SWD.

Assume that H is an SWD. For example, assume that a subject is indifferent between (4,F;6,H) and death in TTO for SWD. We then ask the same subject an L-TTO question; for example, we ask for the value of X such that (10,F;10,H) is as good as (X,F). Which is the value of X such that U(H) is constant with both methods? The answer is that we do not know if we do not know the discount rate. Under the linear QALY model, X should be 3.33 and then U(H) = –2/3. However, if the discount rate were 5% and X = 3.33, we would find that U(H) = –0.85 with TTO and U(H) = –1.00 with L-TTO. We do not know

Table 1 Dependency States: Brief Description of Attributes and Levels

Eating	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 fecal) and needs help for hygiene.
	3. Has both urinary and fecal incontinence and needs help for hygiene.
Personal care	1. Does not need help for personal care: bathing, dressing, etc.
	2. Needs help only to bathe 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 the home.
	3. Needs help to move both in and out of the 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.
Cognition problems	1. Does not need help due to mental or cognitive problems or has no mental or cognitive problems.
	2. Needs assistance to manage money or medication or to make some common everyday decisions. Collaborative attitude with the caretaker.
	3. Incapable of taking basic decisions. Cannot live alone. Does not offer resistance.
	4. Incapable of taking basic decisions. Cannot live alone. Does not collaborate and usually offers resistance to help.

whether U(H) is the same in both methods unless we know the discount rate.

In summary, in the case of SBD, the clearest candidate to explain disparities between TTO and L-TTO is violations of additive separability. Previous literature has provided conflicting evidence about this issue. Devlin and others[7] concluded that L-TTO produced higher values in 4 of the 10 states evaluated for SBD (no differences in the other 6 states). Attema and others[8] found that L-TTO produced lower utilities than TTO in 3 of the 6 states evaluated for SBD (no differences in the other 3 states).

The second objective is to test whether the potential differences between TTO and L-TTO are related to the severity of the heath states. The more severe the health state, the more relevant it could be to know in advance that we will be ill in the future. It will also be tested, in a similar way as in objective 1, whether the probability that a health state is worse than dead depends on the severity of the health state.

METHODS

Selection of Health States

The survey used in this study was funded to estimate utilities for health states associated with different levels of dependency generated by health problems. The descriptive system is shown in Table 1. It gives rise to 1728 possible health states. We applied the method of optimal design to reduce the number of combinations to be evaluated by the participants in the final survey. The OPTEX Procedure from SAS Software (version 9.1) was used to generate a set of 24 health states divided into 4 blocks of 6 (Table 2). Each participant in the survey valued only 1 of 4 blocks (6 health states). Blocks were randomly allocated among subjects. We also randomized order of presentation of health states. Each participant used only TTO or L-TTO; that is, we used a between-samples design. To test our hypotheses, we selected those responses that implied that the health state was better than dead.

Selection of Respondents

Subjects were selected from the general population of Galicia (a region in the northwest of Spain), using a 4-stage cluster stratified random sampling. A total of 500 interviews were conducted: 312 participants responded to the TTO protocol and 188 participants to the L-TTO protocol. Interviews were conducted face to face by trained interviewers. First, we used the TTO protocol to estimate a scoring algorithm for the health state classification system shown. A minimum number of 300 subjects was established for that purpose. Once this objective

Table 2 Dependency States Valued by Block

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

Note: The number indicates the level of each attribute following the order of Table 1.

was achieved, we used the rest of the sample to study methodological issues related with L-TTO.

The questionnaires. Two types of questionnaires were used: one for TTO and another for L-TTO. Both began explaining the objective of the study and the health states (dimensions and levels) used in the questionnaire. Next, subjects had to evaluate 6 health states with TTO or L-TTO. We also asked subjects to state the 2 attributes that were most important for them. Finally, we collected the sociodemographic characteristics of the participants: age, gender, family income, education, labor status, living arrangements, size of municipality, own health (measured by Euroqol EQ-5D), whether they knew a dependent relative, and whether the relative lived with them.

Valuation Procedure

The procedure is illustrated in Figure 1. Subjects had to choose between 2 options (A and B) with different health profiles. Visual aids were used to help the subject understand these questions. The first question classified a health state as better or as worse than dead. In TTO, the first question involved choosing between immediate death and 10 years in a certain health state (H). In L-TTO, the first question involved choosing between 10 years in full health (F) and 10 years in F followed by 10 years in H.

Depending on the answer to the first question, the respondent followed a different path as shown in Figure 1. To clarify the procedure, we give 2 examples, 1 for TTO and 1 for L-TTO:

- TTO: Assume that somebody preferred (10 years, H) to death; then she was asked to choose between (10 years, H) and (5 years, F). The number of years in full health was moved up and down until an indifference interval (or value) was reached. The middle point of the interval was estimated as the indifference point. For example, if (8 years, F) $\succ$ (10 years, H) and (7 years, F) $\prec$ (10 years, H), we assumed that (7.5 years, F) ~ (10 years, H) and 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, F; 10 years, H) to (10 years, F); then she was asked to choose between (15 years, F) and (10 years, F; 10 years, H). Applying the procedure used in TTO, we obtained an indifference (or value) interval. For example, if (10 years, F; 10 years, H) ~ (12 years, F), then 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.

Analysis

Consistency. Violations of dominance were analyzed to test the internal consistency of responses. A health state dominates another if it is at least better in 1 dimension and it is not worse in any of the other dimensions. For instance, in block 1 the health state 313331 dominates 323433 (see Table 2). There are 6 cases of dominance in blocks 1–3 and 4 cases in block 4. We identified the participants who violated dominance at least once. Dominance is violated if a worse health state is valued higher than a better health state.

Hypotheses. To achieve the first objective, the following hypotheses were tested:

Hypothesis 1: Utilities for SBD are not systematically different between TTO and L-TTO.

Hypothesis 2: The probability that a health state is considered worse than dead does not change systematically between TTO and L-TTO.

To test our hypotheses, we first formulated the following model:

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

where U_{ij} is the utility assigned by respondent i (using TTO or L-TTO) to health state j (j = 1, 2, . . . , 24); s_j is a dummy variable that identifies the state valued (e.g., $s_j = 1$ if $j = 1$ and $s_j = 0$ if $j \neq 1$); x_i is a vector of personal characteristics of the participants; *Lead* is a dummy variable indicating whether the participant evaluated the health state using L-TTO (*Lead* = 1) or

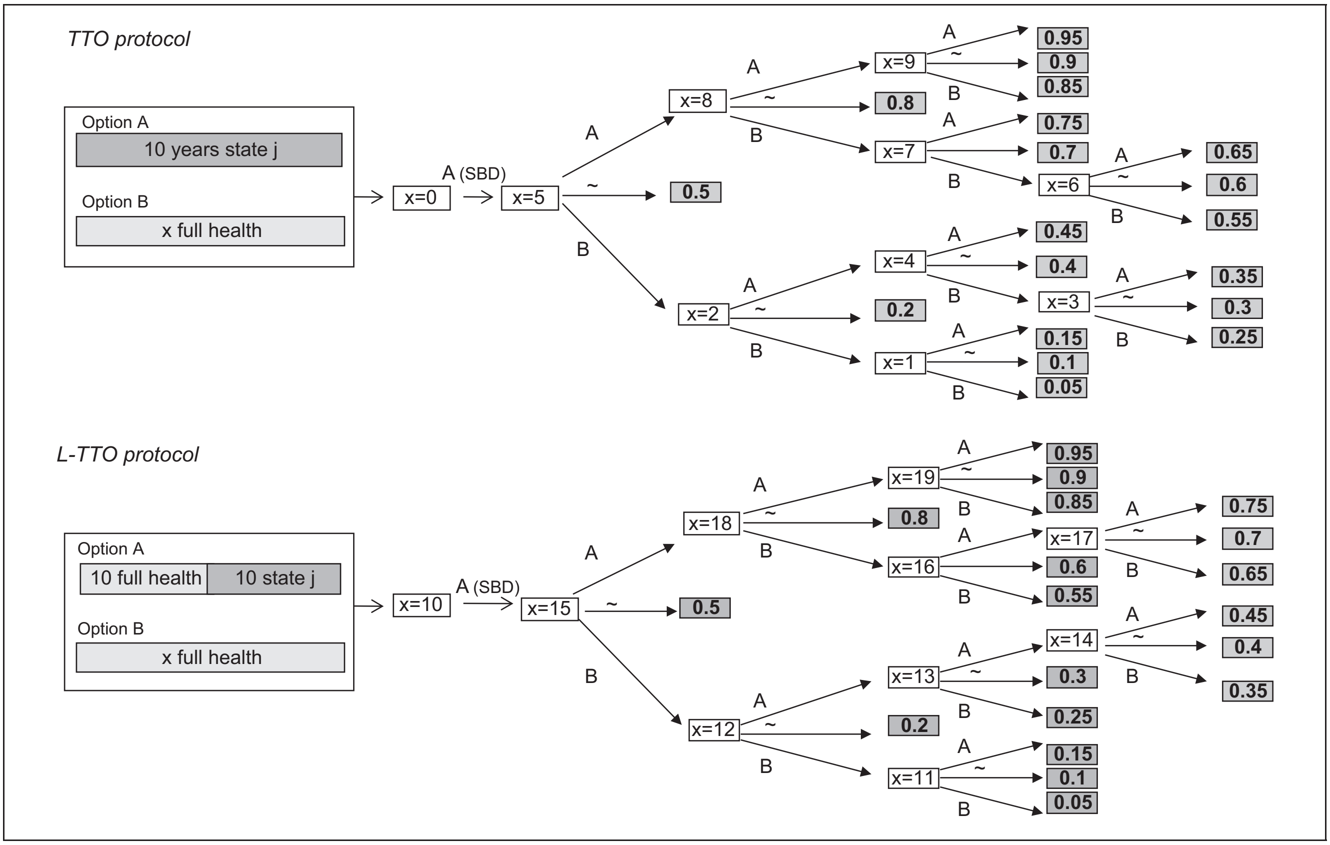

Figure 1 *Protocol used in the questionnaire: time tradeoff (TTO) version and lead time tradeoff (L-TTO) version.*

TTO (*Lead* = 0); ε_{ij} is an error term; and α, β_j, δ', and γ are the parameters to be estimated.

To test hypothesis 1, we restricted U_{ij} to be positive. It was estimated using the random effects regression model since it takes into account that the observations provided by the 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 with each observation. We test hypothesis 1 with the parameter γ. If it is statistically different from zero, we will conclude that L-TTO and TTO produce systematically different utilities. To test hypothesis 2, U_{ij} was transformed into a binary variable taking a value of 1 if the respondent considered this state worse than dead and 0 otherwise. This model was estimated using a random effects logit model to capture unobserved factors specific to each respondent. We test hypothesis 2 with the new parameter γ. In this model, a parameter γ statistically different from zero and with a negative (positive) sign indicates that the probability that a state is considered worse than dead is lower (higher) with L-TTO than with TTO.

The 2 hypotheses were tested using 2 different specifications of the respective model (Stata statistical package was used for all analysis). In one specification sociodemographic variables were included (income was excluded from the analysis because 9.6% of subjects did not respond to this question), and in another one they are excluded.

To achieve the second objective, we tested the following hypotheses:

Hypothesis 3: The differences between the utilities obtained from TTO and L-TTO do not depend on the severity of the states.

Hypothesis 4: The differences between TTO and L-TTO in the probability that a health state is considered worse than dead do not depend on the severity of the states.

The severity of a state was approximated by the proportion of participants who considered this state as worse than dead in TTO. Thus, we considered that a state was more severe than another if a greater percentage of participants found it worse than dead in the TTO method (i.e., they chose option B in the first question). Using this criterion, we ranked the states and classified them into 3 groups of equal size (8 health states in each group), namely *Less severe* (G_1), *Intermediate* (G_2), and *More severe* states (G_3).

To test hypotheses 3 and 4, the following model with interactions was formulated:

$$U_{ij} = \alpha + \sum \beta_j s_j + \gamma_1(G_1 * Lead) + \gamma_2(G_2 * Lead) + \gamma_3(G_3 * Lead) + \varepsilon_{ij}.$$

As in hypothesis 1, to test hypothesis 3 this model was estimated using the random effects regression model, where U_{ij} is the utility assigned by respondent i to health state j, restricting U_{ij} to positive values. If γ_1 is positive (negative) and statistically different from zero, L-TTO produces utilities greater (lower) than TTO for the less severe states (the same interpretation for the rest of the groups). As in hypothesis 2, to test hypothesis 4 we transformed U_{ij} into a binary variable taking a value of 1 if the respondent considered this state worse than dead and 0 otherwise, and then a random effects logit model was estimated. In this estimation, if γ_1 is statistically different from zero and with a negative (positive) sign, then, for the less severe states, the probability that a state is considered worse than dead is lower (greater) with L-TTO than with TTO (the same interpretation for the rest of the groups).

RESULTS

Table 3 shows the characteristics of respondents in both samples. Statistical analysis showed that there were no statistically significant differences (χ^2 test at the 5% level) between the 2 samples except for income and proximity to dependent persons. This shows the relevance of controlling for the characteristics of respondents to test our hypotheses.

Consistency was high. Most respondents never violated dominance (83% in TTO and 72% in L-TTO). Given that subjects can make random errors,[9] and given that there is not a normative criterion to determine when a subject is ''too inconsistent,'' we initially did not exclude anybody from the analysis. We also analyzed our data using only the responses of perfectly consistent subjects in order to check the potential influence of inconsistencies.

The main results are presented in Table 4. Results are consistent also at the aggregate level. We conducted binary comparisons (results not shown) of the parameters of those health states where relations of dominance were observed (22 pairs) using the Wald test. The hypothesis of equality of parameters was rejected in 18 cases (15 at the 1% level and 3 at the 5% level) and always in the expected direction (the parameter of the dominant health state was higher).

The hypothesis that utilities are the same in both procedures (Hypothesis 1) is rejected. The coefficient of the *Lead* variable (γ) is positive and significantly different from zero (models 1 and 2). It is also quite high since L-TTO adds about 0.2 points to the average utility of health states, in relation to the TTO method. This is an important effect if we compare this amount with minimally important differences mentioned in the literature,[10] as it appears in the discussion. Hypothesis 2 cannot be rejected (models 3 and 4). In these models, the coefficient of the *Lead* variable is not significantly different from zero, indicating that the probability that a state is considered worse than dead is not different between TTO and the L-TTO at the aggregate level.

To test Hypotheses 3 and 4 (second objective of the paper), the 24 states were classified in 3 groups of 8 health states as follows: the *Less severe* health states were those that were considered as worse than death by less than 30% of the participants, the *More severe* health states were those that were considered as worse than death by more than 70% of participants, and the rest were the *Intermediate* health states. Both Hypotheses 3 and 4 are rejected. Model 5 in Table 4 shows that the difference between L-TTO and TTO increases with severity. So L-TTO adds about 0.14 points to the average utility of the less severe states, 0.23 points to the intermediate states, and 0.28 points to the more severe states. There are significant differences (Wald test) between the parameters of groups 1 and 2 and of groups 1 and 3 at the 5% level and between groups 2 and 3 at the 10% level. In addition, model 6 shows that the probability that a health state is considered worse than dead is lower with L-TTO for the most severe health states. This result confirms that the disparity between TTO and L-TTO is more important for the most severe health state.

Other auxiliary regressions were conducted (results not shown) to test the stability of the results: (1) sociodemographic variables (with and without income) were included in all models, and (2) we estimated the models including only the participants who verified all dominance tests and including only the participants who did not violate more than 1 test. None of these alternatives models significantly changed the conclusions.

Table 3 Characteristics of Respondents by Type of Questionnaire (%)

	L-TTO (*n* = 188)	TTO (*n* = 312)
Gender		
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 dependent (not living together)	59.0	40.1
Close dependent (living together)	9.6	6.7
Labor 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
Proportion of subjects who considered an attribute as the worst		
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
Proportion of subjects who considered each attribute as the second worst		
Eat	16.5	17.0
Incontinence	45.2	30.1
Personal care	10.1	19.6
Mobility	15.4	20.2
Housework	2.1	2.9
Mental	10.6	10.3

DISCUSSION

The main results of this study are that (a) L-TTO produces significantly higher utilities than TTO for SBD, (b) this effect increases with severity, and (c) the probability that a heath state is considered as worse than dead is not different between TTO and L-TTO, except for the most severe states. We conclude that L-TTO and TTO produce different utilities for SBD. These results suggest that additive separability is violated. More specifically, people perceive that health problems are less severe if a lead period in full health is added upfront.

It is more difficult to say to what extent the mean difference we found (about 0.2 points) is relevant in

Table 4 Results of the Estimation

	Hypothesis 1: Random Regression		Hypothesis 2: Random Logit		Hypothesis 3: Random Regression	Hypothesis 4: Random Logit
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Constant	0.637**	0.646**	–4.596**	–6.544**	0.651**	–4.786**
States (Ref: 211121)						
133334	–0.249**	–0.247**	6.946**	6.935**	–0.315**	7.453**
122222	–0.196**	–0.195**	2.741**	2.718**	–0.195**	2.755**
214232	–0.186**	–0.184**	3.650**	3.619**	–0.215**	3.768**
313331	–0.130**	–0.128**	3.032**	3.006**	–0.161**	3.149**
323433	–0.252**	–0.251**	6.682**	6.665**	–0.324**	7.183**
111221	0.048	0.053	0.218	0.155	0.049	0.173
112132	–0.096**	–0.091**	2.013**	1.965**	–0.095**	1.975**
112211	0.015	0.020	0.218	0.155	0.017	0.173
223234	–0.233**	–0.229**	6.769**	6.709**	–0.300**	7.299**
234333	–0.299**	–0.294**	6.330**	6.274**	–0.368**	6.853**
333122	–0.200**	–0.195**	5.767**	5.716**	–0.231**	5.855**
111112	–0.041	–0.041	1.052	0.916	–0.037	0.981
113233	–0.171**	–0.171**	5.046**	4.927**	–0.202**	5.106**
213322	–0.206**	–0.206**	4.780**	4.661**	–0.237**	4.840**
222131	–0.146**	–0.146**	2.778**	2.655**	–0.144**	2.708**
234431	–0.248**	–0.250**	6.498**	6.390**	–0.280**	6.558**
334234	–0.271**	–0.272**	7.500**	7.407**	–0.330**	8.071**
123121	–0.098**	–0.096**	2.143**	2.200**	–0.096**	2.095**
212223	–0.142**	–0.138**	5.348**	5.427**	–0.176**	5.431**
233432	–0.185**	–0.182**	6.713**	6.790**	–0.247**	7.258**
314434	–0.191**	–0.188**	7.877**	7.954**	–0.258**	8.436**
324332	–0.188**	–0.184**	6.618**	6.695**	–0.242**	7.162**
333231	–0.151**	–0.148**	5.194**	5.274**	–0.187**	5.277**
Lead (γ) (Ref: TTO)	0.179**	0.204**	–0.400	–0.498		
Lead2 (Ref: TTO)						
Less severe group (γ_1)					0.136**	0.230
Intermediate group (γ_2)					0.227**	–0.084
More severe group (γ_3)					0.281**	–1.298**
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 living together		–0.079**		0.350		
Living together		–0.140**		0.068		
Good health (Ref: EQ-5D $\neq$ 11111)		0.036		0.269		
Labor status (Ref: employed)						
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	456	500
Observations	1557	1557	3000	3000	1557	3000

Note: ** and * indicate statistical significance at (respectively) the 5% and 10% levels.

practice. There is also some literature about ''minimally important differences'' for health states utilities.[10–13] This is defined as the difference that is perceived by patients as beneficial or that would result in a change in treatment. It has been estimated that differences of 0.041 in the SF-6D and 0.074 in the EQ-5D represent minimally important differences.[10] In general, differences of 0.03 or more in scores based on generic preference–based measures are considered important.[13] Although these differences have been obtained for generic measures (not for directly obtained TTO scores), we consider that taking these numbers as our reference, 0.2 is important. As Wakker[14] pointed out, QALY model 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.''[(p209)] In our case, the violation of this assumption seems to generate important biases.

One explanation for these results is that the introduction of a lead period in full health allows people to prepare for the bad years that will come. While in TTO the bad years are a surprise (they start immediately), in L-TTO people have time to make adjustments. This explanation is consistent with the fact that differences between methods are greater for more severe states. The more severe the health state, the more important it is to have some time in full health to prepare for those bad years.

There is nothing wrong in violating additive separability. It is a convenient assumption that makes the QALY model more tractable, but it is not a normative assumption. However, some undesirable consequences are derived if it does not hold. One is that it complicates the task of eliciting utilities with L-TTO. Given that additive separability is violated and that it seems to have important consequences in the estimation of utilities, it is necessary to use a multiplicative model or a more general multilinear functional form. It also shows that it is not straightforward to compare TTO and L-TTO utilities since we have to take into account that there are interactions between time periods. This does not mean that we cannot establish a relationship between utilities estimated with TTO and L-TTO. However, the rejection of additive separability implies that utilities estimated with TTO and L-TTO are not easily interchangeable. To estimate utilities for health states with L-TTO, we should use nonadditive models, and then the parameters that reflect the interactions between disjointed time periods need to be estimated separately. Otherwise, the ''utility'' estimated with L-TTO picks up 2 different effects: namely, the severity of the health state and the effect of the interaction between disjointed time periods.

Our design isolates the effect of a constant discount rate, the assumption habitually considered in the literature of the economic evaluation. To test whether the differences between TTO and L-TTO could be eliminated under hyperbolic discounting, all models were again estimated assuming that subjects used hyperbolic discounting. We used 2 models proposed in the literature,[15,16] with the parameters estimated by van der Pol and Cairns[17] in a health context (results can be obtained from the authors on request). However, the main conclusions of our paper remain unaltered; namely, L-TTO produced mean utilities at least 0.15 higher than TTO.

We have considered alternative explanations to our results, such as the effect of loss aversion. The time frame is different in both methods (10 years v. 20 years), and there is evidence[18–20] suggesting that if the time frame used in the study is shorter than subjective life expectancy (as it happens in our case for most people), subjects are reluctant to give up life-years in full health. However, the effect of loss aversion predicts higher utilities for TTO since time horizon is shorter in TTO than in L-TTO.

Let us now compare our results with previous findings. Results reported by Devlin and others[7] are similar to ours. They found that in 4 of the 10 health states analyzed, L-TTO produced higher utilities than TTO for SBD and no differences were found in the others. However, Attema and others[8] reported opposite results; that is, L-TTO produced lower values than TTO in 3 of the 6 health states evaluated and no difference in the rest. Our study is different from the 2 previous studies in several respects. First, we used a wider range of states (24 states). This made it possible to test more accurately whether potential violations of additive separability are related to the severity of the health states. Second, we used a different way of testing our hypotheses. Instead of comparing each health state separately, we conducted regression analysis identifying the method with a dummy. We observed whether the coefficient that identifies the method is significant. This has some advantages. One is that it makes it possible to estimate the overall size of the bias. Another is that for bad health states, the number of observations can be small, since most subjects think they are worse than dead. We may not detect differences if we compare a single health state simply because the sample is small in each of the comparisons. Finally, there are differences in relation to sample composition. Contrary to Attema and others, we used members of the general population as subjects, increasing the external validity of our test. Devlin and others only used L-TTO in their study, whereas we collected data using both methods.

Two reports are not directly comparable to our paper but provide some important evidence about L-TTO. Versteegh and others[21] compared L-TTO with what they call ''lag-TTO,'' where the years in full health come after the years in bad health, finding that L-TTO produced higher utilities than lag-TTO. Devlin and others[22] compared several L-TTO framings and found that the ratio of lead time to disease time had important effects on utilities. This result was also observed by Attema and others[8] and Versteegh and others.[21]

The message that comes from all these papers is that the lead period in full health that precedes the years in bad health influences the perceived severity of those bad years. This evidence adds to what we know from the path-state approach. That is, some investigators[23,24] observed that the score for a series of events when evaluated as a path was not equal to the time-weighted sum of the scores for each of the discrete states that comprised the same path. Krabbe and Bonsel[25] found that subjects were sensitive to the sequence of health states, and the investigators concluded that the utility of a health profile ''may not be regarded as simply a chain of independent separately valued and discounted QALY periods.''[(p178)] All this evidence suggests that additive separability does not seem to reflect preferences well. This calls for a better understanding of these interactions in order to use L-TTO as a standard preference elicitation method. This in itself is not a negative result for L-TTO. It certainly makes the method more elaborate than TTO. However, it also has advantages. For example, in those contexts where illnesses are diagnosed in advance (e.g., Parkinson's or Alzheimer's disease), utilities elicited with L-TTO could reflect preferences better than utilities elicited with TTO since L-TTO will pick up the ''knowing in advance effect.'' Another advantage of L-TTO is that it has been shown to produce utilities with better psychometric properties than TTO for SWD. For example, the method reduces discontinuities around 0 and generates less extreme negative utilities, reducing the problem of aggregating those negative values. Given those advantages over TTO, it is important to better understand how the introduction of the lead period (the main feature of the method) influences the utility of subsequent health states.

This paper is not without limitations. First, we used a between-samples design, and both subsamples were not randomized. A within-sample design has the obvious advantage that it reduces the sources of variability in the response. However, it is more susceptible to anchoring and order effects between methods. If we want to know the difference between using TTO and L-TTO in practice, a between-samples design better reflects what we can find in real life. That is, in practice, researchers are going to use one single method. The disadvantage is that some of the differences between methods may come from differences in preferences. In this context, the best we could do was have similar sociodemographic composition in both subsamples. This was almost achieved, but there were significant differences in some attributes, namely, income and the number of participants with a close dependent. We used multivariate analysis to isolate the influence of these factors. Given that the majority of the characteristics are similar and given that the main results do not change when these characteristics are included or excluded in the regression, the use of a between-samples design and the absence of randomization do not seem to question the results. Second, we used different time horizons (10 years in TTO and 20 years in L-TTO), and this may have introduced a confounding factor in the comparison between both methods. However, empirical evidence suggests[18,26] that the effect of the different time frame may generate the opposite effect that we find, which would reinforce our conclusions. Another limitation is that we did not test additive separability directly in a controlled experiment, as in Treadwell.[27] We attributed to additive separability the difference we found between TTO and L-TTO for SBD. However, the method used by Treadwell has one limitation; namely, it does not reflect how important these violations are in practice. As mentioned above, we know that models are only approximations of reality. Testing whether additive separability holds is, from a practical point of view, less relevant than having an idea of the size of the bias introduced by a model that relies on this assumption. Recently, Devlin and others[7(p348)] asserted 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.

REFERENCES

1. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. Health Serv Res. 1972;7(2):118–33.

2. Torrance GW. Measurement of health state utilities for economic appraisal. J Health Econ. 1986;5(1):1–30.

3. Miyamoto J, Wakker PPZ, Bleichrodt H, Peters H. The zero-condition: a simplifying assumption in QALY measurement. Manag Sci. 1998;44:839–49.

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

5. Tilling C, Devlin N, Tsuchiya A, Buckingham K. Protocols for time trade off valuations of health states worse than dead: a literature review. Med Decis Making. 2010;30(5):610–9.

6. Augestad LA, Rand-Hendriksen K, Sønbø Kristiansen I, Stavem K. Learning effects in time trade-off based valuation of EQ-5D health states. Value Health. 2012;15(2):340–5.

7. Devlin NJ, Tsuchiya A, Buckingham K, Tilling C. A uniform time trade off method for states better and worse than dead: feasibility study of the ''lead time'' approach. Health Econ. 2011; 20(3):348–61.

8. Attema AE, Versteegh MM, Oppe M, Brouwer W, Stolk EA. Lead time TTO: leading to better health state valuations? Health Econ. 2012;22(4):376-392.

9. San Miguel F, Ryan M, Amaya-Amaya M. ''Irrational'' stated preferences: a quantitative and qualitative investigation. Health Econ. 2005;14(3):307–22.

10. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523–32.

11. Wyrwich KW, Bullinger M, Aaronson N, Hays RD, Patrick DL, Symonds T; Clinical Significance Consensus Meeting Group. Estimating clinically significant differences in quality of life outcomes. Qual Life Res. 2005;14(2):285–95.

12. Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health Qual Life Outcomes. 2007;5:70.

13. Feeny D, Spritzer K, Hays RD, et al. Agreement about identifying patients who change over time: cautionary results in cataract and heart failure patients. Med Decis Making. 2012;32(2):273–86.

14. Wakker P. A criticism of healthy-years equivalents. Med Decis Making. 1996;16(3):207–14.

15. Harvey CM. Value functions for infinite period planning. Manag Sci. 1986;32:1123–39.

16. Mazur JE. 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): Erlbaum; 1987.

17. van der Pol M, Cairns J. A comparison of the discounted utility model and hyperbolic discounting models in the case of social and private intertemporal preferences for health. J Econ Behav Org. 2002;49:79–96.

18. van Nooten FE, Koolman X, Brouwer WB. The influence of subjective life expectancy on health state valuations using a 10 year TTO. Health Econ. 2009;18(5):549–58.

19. van Nooten FE, Koolman X, Busschbach JJ, Brouwer WB. Thirty down, only ten to go?! Awareness and influence of a 10-year time frame in TTO. Qual Life Res. 2014;23(2):377–84.

20. Heintz E, Krol M, Levin LA. The impact of patients' subjective life expectancy on time tradeoff valuations. Med Decis Making. 2013;33(2):261–70.

21. Versteegh MM, Attema AE, Oppe M, Devlin NJ, Stolk EA. Time to tweak the TTO. But how? Eur J Health Econ. 2013;14:43–51.

22. Devlin N, Buckingham K, Shah K, et al. A comparison of alternative variants of the lead and lag time TTO. Health Econ. 2013; 22(5):517–32.

23. Richardson J, Hall J, Salkeld G. The measurement of utility in multiphase health states. Int J Technol Assess Health Care. 1996; 12(1):151–62.

24. Kuppermann M, Shiboski S, Feeny D, Elkin E, Washington E. Can preference scores for discrete states be used to derive preference scores for an entire path of events? An application to prenatal diagnosis. Med Decis Making. 1997;17(1):42–55.

25. Krabbe PFM, Bonsel GJ. Sequence effects, health profiles, and the QALY model. Med Decis Making. 1998;18(2):178–86.

26. Lin MR, Yu WY, Wang SC. Examination of assumptions in using time tradeoff and standard gamble utilities in individuals with spinal cord injury. Arch Phys Med Rehabil. 2012;93(2):245–52.

27. Treadwell JR. Tests of preferential independence in the QALY model. Med Decis Making. 1998;18(4):418–28.