Centre de Referència en Economia Analítica

Barcelona Economics Working Paper Series

Working Paper nº 113

The Validity of QALYs Under Non-Expected Utility

Han Bleichrodt and Jose Luis Pinto

February, 2004

The Validity of QALYs Under Non-Expected Utility

Han Bleichrodt, Erasmus University, Rotterdam, The Netherlands

Jose Luis Pinto, Universitat Pompeu Fabra, Barcelona, Spain

Barcelona Economics WP nº 113

February 2004

Abstract

This paper examines applications of non-expected utility in the health domain. The most widely used utility model in health economics, the time-linear QALY model, assumes (i) separability of quality of life and life duration, and (ii) linearity of the utility for life duration. We perform new tests, which are robust to violations of expected utility, of these two assumptions. The data support separability, but show that the utility for life duration is concave rather than linear. The finding of concave utility may not be surprising in itself. The contribution of this paper is to demonstrate this empirically without being invalidated by violations of expected utility.

KEY WORDS: Non-expected Utility, QALYs, Utility curvature, Aggregation.
JEL CLASSIFICATION: D81, I10

ADDRESS CORRESPONDENCE TO: Han Bleichrodt, iMTA/iBMG, Erasmus University, PO Box 1738, 3000 DR Rotterdam. +31 10 408 85 66 (Ph). +31 10 408 90 92 (F).
Email: bleichrodt@bmg.eur.nl

1. Introduction.

It has by now been widely recognized that expected utility is not valid as a descriptive theory of decision under uncertainty. The descriptive violations of expected utility have led to the emergence of several non-expected utility theories. The increasing importance of non-expected utility makes it necessary to reassess applications that were previously based on expected utility. Examples of such reassessments include Karni and Safra (1989), Crawford (1990), and Dekel *et al.* (1991) for game theory, Machina (1995) and Wakker *et al.* (1997) for insurance theory, and Cubitt and Sugden (1998) for the evolution of preferences. See Starmer (2000) for a review. The present paper examines applications of non-expected utility in the health domain. We focus on the main utility model in health economics, the quality-adjusted life-years (QALY) model, and present new theoretical foundations for, and empirical tests of, this model under non-expected utility.

QALYs provide a simple way to combine the two dimensions of health, life duration and health status, into a single utility index. They are intuitively appealing, which facilitates communication to policy makers, and analytically tractable, which explains their widespread use in practical studies. A disadvantage of QALYs is that they only represent individual preferences over health under strong assumptions. If these assumptions do not hold then the use of QALYs may lead to incorrect policy recommendations. To gain insight into the validity of QALYs, it is necessary to assess the restrictiveness of the assumptions that the QALY model imposes.

In the most common version of the QALY model, the time-linear QALY model, the utility of the health outcome of spending T years in health state Q is equal to T*V(Q), where V is a utility function over health states. That is, in the time-linear QALY model the utility for duration is linear. It is well known that under expected utility, linearity of utility implies risk neutrality. As it turns out, risk neutrality with respect to life duration is not only necessary, but also sufficient for the time-linear QALY model to correctly represent individual preferences for health under expected utility (Bleichrodt *et al.*, 1997, Miyamoto *et al.*, 1998). Empirical tests of risk neutrality with respect to life duration typically yielded negative results: people are not neutral, but averse, towards duration risk (McNeil *et al.*, 1978, McNeil *et al.*, 1981, Stiggelbout *et al.*, 1994, Verhoef *et al.*, 1994). Stiggelbout *et al.* (1994), in a sample of testicular cancer patients, found, for example, that their median respondent was indifferent between 4 years in good health for sure and a treatment giving a probability ½ of 10 years in good health and a probability ½ of death, i.e. 0 years in good health. Under expected utility, risk aversion with respect to duration is incompatible with linear utility for life duration and the above studies, therefore, suggest that the time-linear QALY model should be rejected.

Many studies have shown that people behave in ways that systematically violate expected utility; for an overview see Camerer (1995) and Starmer (2000). Given that expected utility does not hold, the above tests of the time-linear QALY model, which are based on expected utility, are inconclusive. Under non-expected utility, people can both be risk averse with respect to life duration and have linear utility for duration. Suppose, for example, that an individual maximizes rank-dependent utility (Quiggin, 1981) and consider, again, the median respondent in the study by Stiggelbout *et al.* (1994). Under rank-dependent utility the observed indifference implies that U (4 years in good health) = w(½)*U(10 years in good health) + (1−w(½))*U(0 years in good health), where w is a probability weighting function that satisfies w(0) = 0 and w(1) = 1. It is easy to verify that, under rank-dependent utility, the median indifference in the study by Stiggelbout *et al*. is consistent with linear utility for life duration if w(½) = 0.4.

Because risk aversion with respect to life duration does not necessarily exclude linear utility for life duration under non-expected utility, the question of whether the utility for life duration is linear is still open. This is unfortunate given the importance of the time-linear QALY model in health economics and medical decision making. The aim of this paper is, therefore, to develop and perform new tests of the descriptive validity of the time-linear QALY model that are robust to violations of expected utility.

The first part of the paper is theoretical and derives the conditions that are critical in two QALY models under non-expected utility. Using some recent results of Bleichrodt and Miyamoto (2003), we first present a preference foundation for the time-linear QALY model and then give a preference foundation for a more general QALY model in which the utility function for duration can be curved. We refer to this latter model as the time-nonlinear QALY model. As a corollary to our representation theorems, we obtain a new characterization, that is more flexible in applications than the existing characterizations, of Choquet expected utility (Schmeidler, 1989), currently the main descriptive theory for decision under uncertainty.

The second part of the paper is empirical and tests, by means of two experiments, the critical conditions that were identified in the first part of the paper. The experimental data violate the time-linear QALY model, but generally support the time-nonlinear QALY model. This is an important finding for practical research, because the time-nonlinear QALY model, while less parsimonious than the linear QALY model, is still tractable.

The structure of the paper is as follows. Section 2 presents notation and assumptions. Section 3 gives preference foundations for the time-linear and the time-nonlinear QALY model. Section 4 derives empirical tests of the critical conditions of the time-linear and the time-nonlinear QALY models. Section 5 describes the design and the results of the two experiments that performed these empirical tests. Section 6 concludes. Proofs, extensions, and experimental details are available in the appendices.

2. Notation and assumptions

We consider an individual in a situation where there are two alternative *states of nature*, r and s, exactly one of which pertains. The states of nature can, for example, describe the results of a medical treatment with r and s referring to two mutually exclusive diseases. We consider decision under uncertainty where probabilities for the two states of nature may, but need not be known. The restriction to two states of nature is made for expositional purposes. The generalization to an arbitrary finite number of states of nature is available in Appendix C.

The individual's problem is to choose between *acts*. Each act is a pair of *outcomes*, one for each state of nature. We shall write $f = (f_r, f_s)$ for the act which yields f_r if state of nature r occurs and f_s if state of nature s occurs. An act is *constant* if $f_r = f_s$. In our application, the outcomes are chronic health states, i.e., pairs (q,t) denoting t years in health state q. We write $\mathcal{H}$ for the set of chronic health states and, for notational convenience, we denote outcomes as x,y instead of (q,t),(q′,t′), if no confusion can arise. The life durations t lie in an interval $\mathcal{T} = [0,\mathcal{M}]$, where $\mathcal{M}$ denotes the maximum life duration. In many applications, the set of health states is considered finite and not a continuum. We, therefore, impose no assumptions on the health states $\mathcal{Q}$.

The conventional notation $\succ$, $\succcurlyeq$, and ~ is used to denote relations of strict preference, weak preference, and indifference. We assume that $\succcurlyeq$ is transitive and that for all acts f and g, either $f \succcurlyeq g$ or $g \succcurlyeq f$. Preferences over outcomes are derived from preferences over constant acts, i.e. $x \succcurlyeq y$ if $(x,x) \succcurlyeq (y,y)$. We assume that all acts are *rank-ordered*, that is, the outcome under state of nature r is always weakly preferred to the outcome under state of nature s ($f_r \succcurlyeq f_s$). We denote the set of acts by $\mathcal{H}^2_{\downarrow}$, where the downward arrow serves as a reminder that the acts are rank-ordered. Throughout the paper, statements of the form "for all acts f…" (for all outcomes x, for all durations t, for all health states q) should be read as " for all acts f in $\mathcal{H}^2_{\downarrow}$ …" (x in $\mathcal{H}$, t in $\mathcal{T}$, q in $\mathcal{Q}$).

We assume that in any health state the individual prefers more life duration to less. That is, the preference relation $\succcurlyeq$ satisfies *monotonicity in duration*: for all chronic health states (q,t), (q,t′) with t > t′, (q,t) $\succ$ (q,t′). Because health status is, typically, not quantitative, we cannot define monotonicity with respect to health status. We assume instead that health status is preferentially independent. Preferential independence means that preferences over health states, with duration kept fixed, are independent of the value at which life duration is kept fixed. Formally, health status is *preferentially independent* if for all life durations t,t′ unequal to zero and for all health states q,q′, (q,t) $\succcurlyeq$ (q′,t) $\Leftrightarrow$ (q,t′) $\succcurlyeq$ (q′,t′). To avoid triviality, we assume that not all health states are equivalent: there exist chronic health states (q,t), (q′,t) such that (q,t) $\succ$ (q′,t).

To obtain maximal generality of the tests that we will derive and perform, we did not select one of the existing theories of decision under uncertainty as the framework for our investigations, but assumed, instead, that preferences over acts can be represented by the following general decision rule:

$$f \succcurlyeq g \Leftrightarrow U_r(f_r) + U_s(f_s) \geq U_r(g_r) + U_s(g_s) \qquad (1)$$

where the functions U_r and U_s assign a real-valued index to every chronic health state in $\mathcal{H}$. The functions U_r and U_s are state-dependent and need not be the same. Miyamoto and Wakker (1996) showed that Expression (1) has as special cases expected utility (the case where $U_r(f_r) = p_rU(f_r)$ and $U_s(f_s) = p_sU(f_s)$) and several non-expected utility theories, including influential theories as rank dependent utility (Quiggin, 1981, Yaari, 1987), Choquet expected utility (Schmeidler, 1989), state-dependent expected utility (Karni, 1985), disappointment aversion theory (Gul, 1991), and prospect theory (Kahneman and Tversky, 1979, Tversky and Kahneman, 1992) for gains and losses separately. The results that we will derive in the remainder of the paper are, therefore, valid under all the aforementioned theories and are robust to the deviations from expected utility modeled by these theories.

We assume that U_r and U_s both agree with the preference relation over outcomes. That is, for all chronic health states x,y,

$$x \succcurlyeq y \Leftrightarrow U_r(x) \geq U_r(y) \Leftrightarrow U_s(x) \geq U_s(y).$$

We do not assume, however, that U_r and U_s order utility differences the same way. Consequently, $U_r(x) - U_r(y) \geq U_r(x') - U_r(y')$ and $U_s(x) - U_s(y) < U_s(x') - U_s(y')$ can occur simultaneously. If U_r and U_s order utility differences the same way, then they must be *linear with respect to each other*: there exist positive $\sigma > 0$ and real τ such that $U_r = \sigma U_s + \tau$. If U_r and U_s are linear with respect to each other then, given that τ can be chosen 0 while maintaining Expression (1), they can be chosen equal to $\pi_r U$ and $\pi_s U$, respectively, where π_r and π_s are positive decision weights, e.g. subjective probabilities, and U is a real-valued utility function on $\mathcal{H}$ (Miyamoto and Wakker, 1996). Linearity of U_r and U_s with respect to each other will follow from the conditions that we impose to characterize the QALY models.

Wakker (1993) gave a preference foundation for Expression (1) for the case where the outcome set is a continuum. Wakker's preference axioms are in Appendix A. Because we made no assumptions about the set of health states $\mathcal{Q}$, our outcome set $\mathcal{H}$ is not necessarily a continuum, and, therefore, Wakker's proof does not apply in the decision context of this paper. Using two results from Bleichrodt and Miyamoto (2003), it is straightforward, however, to extend Wakker's result to a domain that is not a continuum if the *zero-condition* holds. This condition asserts that for a life duration of zero all health states are equivalent: for all health states q,q′, (q,0) ~ (q′,0). The condition is self-evident in the medical context of this paper, because (q,0) and (q′,0) are indistinguishable under the interpretation of time as life duration (Miyamoto and Eraker, 1988, Bleichrodt *et al.*, 1997, Miyamoto *et al.*, 1998). This extension of Wakker's result to outcome sets that are not a continuum is presented in Appendix B.

3. QALY Characterizations

The Time-Linear QALY Model

The *time-linear QALY model* holds if, in Expression (1), $U_r = \pi_r U$ and $U_s = \pi_s U$, where π_r and π_s are positive decision weights and U is a real-valued function on $\mathcal{H}$, and, moreover, $U = V(q)\cdot t$ with V a health utility function that assigns a positive index to every health state in $\mathcal{Q}$. To characterize the time-linear QALY model, we must find conditions that ensure that the utility functions U_r and U_s order utility differences the same way and that the resulting common utility function U is linear in duration. Remarkably, we can achieve both goals with one single condition, constant marginal utility for life-years, or constant marginal utility for short.

It is well known that a utility function U is linear in life duration if the marginal utility of life-years is constant. That is, for all health states q, and for all life durations $t_1,t_2,t_1+\tau,t_2+\tau$, we have $U(q,t_1+\tau) - U(q,t_2+\tau) = U(q,t_1) - U(q,t_2)$. To be able to express constant marginal utility in terms of the preference relation $\succcurlyeq$, so that it becomes directly testable, we introduce a new definition. We define $[t_1;t_2] \succcurlyeq^* [t_3;t_4]$ if there exists a health state q such that either

$$\big((q,t_1),(q,t'_s)\big) \succcurlyeq \big((q,t_2),(q,t''_s)\big) \text{ and}$$

$$\big((q,t_3),(q,\, t'_s)\big) \preccurlyeq \big((q,t_4),(q,t''_s)\big)$$

or

$$\big((q,t'_r),(q,t_1)\big) \succcurlyeq \big((q,\, t''_r),(q,t_2)\big) \text{ and}$$

$$\big((q,\, t'_r),(q,t_3)\big) \preccurlyeq \big((q,\, t''_r),(q,t_4)\big).$$

when all acts involved are in $\mathcal{H}^2_{\downarrow}$. We define $[t_1;t_2] \succ^* [t_3;t_4]$ if at least one of the above preferences is strict.

It can be shown under Expression (1) that $[t_1;t_2] \succcurlyeq^* [t_3;t_4]$ implies either $U_r(q,t_1) - U_r(q,t_2) \geq U_r(q,t_3) - U_r(q,t_4)$ or $U_s(q,t_1) - U_s(q,t_2) \geq U_s(q,t_3) - U_s(q,t_4)$ for some health state q. Under Expression (1), the $\succcurlyeq^*$ relation can, therefore, be interpreted as measuring utility differences.

With the aid of the $\succcurlyeq^*$ relation, we can translate constant marginal utility into a testable preference condition. We say that *constant marginal utility for life-years* holds if for all life durations t_1, t_2,$t_1+\tau$,$t_2+\tau$, $[t_1+\tau,t_2+\tau] \succ^* [t_1,t_2]$ is excluded. Constant marginal utility implies that U_r and U_s are both linear in life duration for all q and, therefore, order utility differences the same way. We are now in a position to state our first result, which extends Theorem 2 in Bleichrodt and Quiggin (1997) to outcome sets that are not a continuum. The result identifies the critical empirical test of the time-linear QALY model also if expected utility does not hold.

Theorem 3.1.

Under the assumptions made in Section 2 and the zero-condition, the following two statements are equivalent.

(i) Constant marginal utility for life-years holds.

(ii) The time-linear QALY model holds.

□

A proof of Theorem 3.1 is in Appendix D.

The Time-Nonlinear QALY Model

The *time-nonlinear QALY model* holds if in Expression (1), $U_r = \pi_r U$ and $U_s = \pi_s U$, where π_r and π_s are positive decision weights and U is a real-valued function on $\mathcal{H}$ and, moreover, $U = V(q) \cdot W(t)$ where $V: \mathcal{Q} \to \mathbb{R}^+$ is a positively-valued health utility function and $W: \mathcal{T} \to \mathbb{R}$ is a real-valued, strictly increasing, and continuous utility function over life duration. The time-nonlinear QALY model has two characteristic properties. First, utility is independent of the state of nature and, second, the utility of life duration is independent of health status. To characterize and critically test the time-nonlinear QALY model, we must, therefore, find a condition that implies these two properties. We cannot use constant marginal utility because, as we saw in Theorem 3.1, this condition implies that utility is linear in life duration and we want to leave open the possibility that utility is curved. The condition we use to characterize the time-nonlinear QALY model is utility independence of life duration, called utility independence for short. Utility independence says that if health status is kept fixed at a particular level then preferences are independent of the level at which health status is kept fixed. Formally, life duration is *utility independent* on $\mathcal{H}^2_\downarrow$ if $\big((q,t_1), (q,t_2)\big) \succcurlyeq \big((q,t_3), (q,t_4)\big) \Leftrightarrow \big((q',t_1), (q',t_2)\big) \succcurlyeq \big((q',t_3), (q',t_4)\big)$. Utility independence is widely used in decision analysis where it is assumed to hold for all acts. Here we modify the common definition by requiring it for rank-ordered acts only.

Theorem 3.2.

Under the assumptions made in Section 2 and the zero-condition, the following two statements are equivalent.

(i) Life duration is utility independent on $\mathcal{H}_{\downarrow}^{2}$.

(ii) The time-nonlinear QALY model holds.

□

A proof of Theorem 3.2, which uses a technique developed by Miyamoto and Wakker (1996), is available in Appendix D.

Theorems 3.1 and 3.2 show that the assumptions of constant marginal utility and utility independence, respectively, imply that U_r and U_s can be decomposed into a state-dependent decision weight and a state-independent utility function. The resulting model is, in fact, a Choquet expected utility functional (Schmeidler, 1989), a point that we will not elaborate on here. Theorems 3.1 and 3.2 show that the characterizations of the two QALY models give a preference foundation for Choquet expected utility "free of charge" so to say, i.e., without the need to impose additional assumptions.

Previous characterizations of Choquet expected utility imposed "richness conditions": either the set of states of nature was assumed to be infinitely large or the outcome domain was assumed to be a continuum. These richness conditions are not always fulfilled in practical applications. In environmental and health decisions, they are, for example, unlikely to hold. Because we do not impose such richness conditions, our characterization of Choquet expected utility may be more useful in applications.

4. Design of the Empirical Tests of Constant Marginal Utility and Utility Independence

Theorems 3.1 and 3.2 show that the validity of the time-linear and the time-nonlinear QALY model hinge on the validity of constant marginal utility and utility independence, respectively. Both conditions impose restrictions on the utility function for life duration. To test these conditions we performed two experiments in which we elicited utility functions for life duration for each subject and examined whether these conditions were fulfilled.

A problem in utility measurement is that the common elicitation techniques assume expected utility and are, consequently, sensitive to violations of expected utility. The trade-off method was developed by Wakker and Deneffe (1996) to measure utilities when people do not evaluate probabilities linearly, as in expected utility, but transform probabilities. The trade-off method can also be used to elicit the functions U_r and U_s in Expression (1), as we will show below. This means that the utilities elicited by the trade-off method are insensitive to the violations of expected utility modeled by the theories that are consistent with Expression (1) and, therefore, that our tests of constant marginal utility and utility independence are not affected by these violations either.

Another advantage of the trade-off method is that the method that is used to measure utility empirically is the same as the method that is used to axiomatize the model. This unity makes it possible to test models directly by looking at utility measurements. In the next two subsections, we show how the measurements by the trade-off method can be used to assess the validity of the time-linear QALY model and the time-nonlinear QALY model.

The empirical findings in Wakker and Deneffe (1996) suggest that constant marginal utility need not hold. They did not perform statistical tests of constant marginal utility, however. Our tests of utility independence are new.

4.1. Elicitation of the Utility Function for Life Duration

The first step in the trade-off method is to specify two states of nature r and s, two "gauge life durations" M and m, a starting outcome t_0, and a health state q. Because health status is kept fixed during the elicitation of the utility function for life duration, we denote, for notational convenience, outcomes (q,t) as t throughout this subsection. In our experiments, we selected M = 55 years, m = 45 years, and t_0 = 0 years. The description of the states of nature and the selected health states is given in Section 5.

The first question in the trade-off method asks a subject to specify the life duration $t_{1,q}$ so that he is indifferent between (55, 0) and (45, $t_{1,q}$). Recall that the notation (55, 0) means 55 years (in health state q) if state of nature r obtains and 0 years (in health state q) if state of nature s obtains. The subscript q in $t_{1,q}$ serves as a reminder that the elicited duration will, in general, depend on the level at which health status is kept fixed.

If $t_{1,q} \leq 45$ then both acts (55, 0) and (45, $t_{1,q}$) are rank-ordered and Expression (1) implies that

$$\begin{aligned} &(55, 0) \sim (45, t_{1,q}) \\ \Leftrightarrow\ &U_r(55) + U_s(0) = U_r(45) + U_s(t_{1,q}) \\ \Leftrightarrow\ &U_s(t_{1,q}) - U_s(0) = U_r(55) - U_r(45) \end{aligned} \tag{2}$$

After the elicitation of $t_{1,q}$, the subject was asked for the life duration $t_{2,q}$ that made him indifferent between (55, $t_{1,q}$) and (45, $t_{2,q}$). If $t_{2,q} \leq 45$ then both acts are rank-ordered and Expression (1) implies that

$$\begin{aligned} &(55, t_{1,q}) \sim (45, t_{2,q}) \\ \Leftrightarrow\ &U_r(55) + U_s(t_{1,q}) = U_r(45) + U_s(t_{2,q}) \\ \Leftrightarrow\ &U_s(t_{2,q}) - U_s(t_{1,q}) = U_r(55) - U_r(45) \end{aligned} \tag{3}$$

A comparison between (2) and (3) shows that

$$U_s(t_{2,q}) - U_s(t_{1,q}) = U_s(t_{1,q}) - U_s(0) \tag{4}$$

That is, the utility difference between $t_{2,q}$ and $t_{1,q}$ is equal to the utility difference between $t_{1,q}$ and $0 = t_{0,q}$ when the evaluation is performed in terms of U_s.

We can proceed in the above fashion and elicit life durations $t_{j,q}$ for which the subject is indifferent between $(55, t_{j-1,q})$ and $(45, t_{j,q})$. As long as $t_{j,q} \le 45$, this procedure leads to a sequence of durations $\{t_{1,q},\ldots, t_{k,q}\}$ for which $U_s(t_{i,q}) - U_s(t_{i-1,q}) = U_s(t_{j,q}) - U_s(t_{j-1,q})$ with $1 \le i,j \le k$. The function U_s is unique up to origin and unit (see Appendix A) and we can, therefore, scale U_s such that $U_s(0) = 0$ and $U_s(t_{k,q}) = 1$. It then follows for all $0 \le j \le k$ that $U_s(t_{j,q}) = \frac{j}{k}$.

Note that it is crucial that for all j, $t_{j,q} \le 45$. If this condition does not hold then $(45,t_{j,q})$ is not rank ordered and the above analysis is not valid.

4.2. Test of Constant Marginal Utility

By Theorem 3.1, constant marginal utility implies that the utility for duration is linear. Recall that the elicited sequence $\{t_{1,q},\ldots, t_{k,q}\}$has the property that $U_s(t_{i,q}) - U_s(t_{i-1,q}) = U_s(t_{j,q}) - U_s(t_{j-1,q})$ for $1 \le i,j \le k$. Hence, if we find that the difference between successive elements of the sequence $\{t_{0,q}, t_{1,q},\ldots, t_{k,q}\}$, the *step size*, is constant, then this implies that U_s is linear in life duration. It does not mean, however, that constant marginal utility holds because U_r and U_s can be different. A full test of constant marginal utility would require the assessment of two sequences, one in terms of U_r and one in terms of U_s, and the verification that the step size is constant in both of these sequences. In the experiments descibed in Section 5, we only elicited a sequence in terms of U_s. The results in Wakker and Deneffe (1996) suggest that constant marginal utility does not hold. We found violations of constant marginal utility in pilot sessions we performed prior to the actual experiment. Because of these findings, we expected to observe violations of constant marginal utility. To reduce the cognitive burden for the subjects, we, therefore, decided to elicit only one sequence per subject.

4.3. Test of Utility Independence

As noted in Section 3, utility independence asserts that if health status is kept fixed at a particular level then preferences are independent of the level at which health status is kept fixed. This means that the elicited sequence $\{t_{1,q},\ldots, t_{k,q}\}$ should be independent of q. To test for utility independence, we, therefore, elicited sequences $\{t_{1,q},\ldots, t_{k,q}\}$ for different health states q and tested whether they were equal by comparing their step sizes.

5. Experiments

5.1. First Experiment

Because of the importance of the time-linear QALY model in health economics, the aim of our first experiment was to try to replicate Wakker and Deneffe's (1996) findings on constant marginal utility, using a different experimental design. The differences in experimental design between our study and Wakker and Deneffe are described in the next paragraphs. Contrary to Wakker and Deneffe (1996), we also performed statistical tests of constant marginal utility. The data of the first experiment were also used, in combination with those from the second experiment, to test utility independence. Before administering the actual experiment, the experimental design was first tested in several pilot sessions using university staff as subjects.

Fifty-one economics students at the University Pompeu Fabra, Barcelona participated in the experiment. They were paid €36. Responses were elicited in personal interview sessions, which is contrary to Wakker and Deneffe who used group sessions. Personal interviews were chosen to increase the quality of the data. The use of students as subjects limits the generalizability of our findings. Empirical evidence on health utility measurement suggests, however, no systematic differences in the patterns of responses obtained using convenience samples and those obtained using representative samples from the general population. For a review see de Wit *et al.* (2000).

To motivate subjects, we started the experiment by explaining why it is important for health policy to obtain insight into how people value health states. The subjects were then told to imagine that they suffer from a health problem and that the symptoms they display indicate one of two possible diseases. To avoid potential framing effects, the diseases were left unspecified and were labeled A and B. Subjects were told that it is known from medical experience that half of the people with these symptoms contract disease A and the other half contract disease B. The axiomatic analysis presented in Sections 2 and 3 was performed for decision under uncertainty, i.e., without the need to specify probabilities. In the empirical analysis we decided to specify the probabilities, because the pilot sessions showed that subjects found it easier to make tradeoffs when they had explicit information about probabilities. We selected a probability of one half, because this is the most familiar probability. In contrast, Wakker and Deneffe (1996) did not specify the probabilities.

Subjects were told that there exist two treatments to fight the diseases, but that the effectiveness of the treatments depends on which disease they actually have. To be effective, treatment has to start immediately, that is, before the actual disease is known. A translation of the questionnaire is available in Appendix E.

Because we elicited preferences over health, the outcomes in our study had to be hypothetical. Several studies have addressed the question whether response patterns differ between questions with hypothetical outcomes and questions with real outcomes (see Hertwig and Ortmann (2001) for an extensive review). These studies used moderate monetary amounts as outcomes. The general conclusion from these studies is that the effect of real incentives varies across decision tasks. For the kind of taks that we asked our subjects to perform, there appears to be no systematic difference in the general pattern of responses, although real incentivies tended to reduce data variability.

Subjects started with a practice question to familiarize them with the trade-off method. They were asked to explain their answer to the practice question. This explanation allowed us to check whether subjects understood the decision problem and the trade-off method. Once we were convinced that they understood these, we moved on to the actual experiment.

Health status was kept fixed at good health (gh), i.e., no health impairments. We asked each subject 6 trade-off questions, i.e., we elicited for each subject a sequence $\{t_{1,gh},\ldots,t_{6,gh}\}$. We had learnt from the pilot sessions that people find the trade-off method easier to answer if they first determine the life durations for which they consider one of the treatments clearly superior and then move towards their indifference value. We, therefore, first asked subjects to compare the treatments (55, $t_{j-1,gh}$) and (45, $t_{j,gh}$) for $t_{j,gh} = t_{j-1,gh}$ and for $t_{j,gh}$ = 45 years, j = 1,…,6. All subjects agreed that the treatment (55, $t_{j-1,gh}$) is better than the treatment (45, $t_{j-1,gh}$) and all but one that (45, 45) is better than (55, $t_{j-1,gh}$), j = 1,…,6.[1] Subjects were then told that these preferences imply that there should be a value of $t_{j,gh}$ between $t_{j-1,gh}$ and 45 for which their preferences between the treatments switch. Subjects were asked to determine this "switching value" by gradually increasing $t_{j,gh}$ from $t_{j-1,gh}$ and by gradually decreasing $t_{j,gh}$ from 45 until they arrived at a range of values for which they found it hard to choose between the treatments. Subjects were then asked to pick the value of $t_{j,gh}$ for which they considered the treatments most finely balanced from the range of values for which they found it hard to choose.This value was taken as their indifference value $t_{j,gh}$. In contrast with our elicitation

[1] This subject preferred (55,$t_{4,gh}$) to (45,45). To reach indifference $t_{5,gh}$ had to exceed 45. In consequence, the act (45,$t_{5,gh}$) was not rank-ordered and the analysis of Section 4 does not hold. This subject was, therefore, excluded.

procedure, Wakker and Deneffe (1996) directly asked respondents to state their indifference value.

Results

Besides the subject described above, one more was excluded from the analyses because he refused to make any tradeoffs. Figure 1 shows the utility function for years in good health based on the median responses. The crosses indicate the median values of the elicited sequence $\{t_{1,gh,\ldots,}t_{6,gh}\}$. The function appears to be concave rather than linear. The null hypothesis that the step sizes are all equal is rejected both by analysis of variance ($p < 0.001$) and by the nonparametric Friedman test ($p < 0.001$). Analysis based on the mean values of $\{t_{1,gh,\ldots,}t_{6,gh}\}$ leads to the same conclusion. Hence, constant marginal utility and, by implication, the linear QALY model are rejected at the aggregate level.

{insert Figure 1 here}

The individual data confirm the conclusions drawn from the aggregate analysis. Let Δ_{j-1}^{j} denote the difference between two successive step sizes of the elicited sequence $\{t_{1,gh},\ldots,t_{6,gh}\}$: $\Delta_{j-1}^{j} = (t_{j,gh} - t_{j-1,gh}) - (t_{j-1,gh} - t_{j-2,gh})$, $j = 2,\ldots,6$. It is easy to verify that positive Δ_{j-1}^{j} corresponds to concave utility for life duration, zero Δ_{j-1}^{j} corresponds to linear utility for life duration, and negative Δ_{j-1}^{j} corresponds to convex utility for life duration. For each subject, we observed 5 values of Δ_{j-1}^{j}. To account for response error, we classified a subject's utility function for life duration as concave if at least 3 values of Δ_{j-1}^{j} were positive, as linear if at least 3 values of Δ_{j-1}^{j} were zero, and as convex if at least 3 values of Δ_{j-1}^{j} were negative.

Table 1 shows that, even though for some subjects the utility function for life duration is linear, the majority of the subjects have a concave utility function for life duration.

Table 1: Classification of Subjects According to the Shape of Their Utility Function

Shape		
Concave	Linear	Convex
59.2%	26.5%	2.0%

5.2. Second Experiment

The second experiment tested both constant marginal utility and utility independence. To test the robustness of our findings on constant marginal utility, we made two changes in the experimental design in comparison with the first experiment. First, we used other health states than good health. Second, we used a different method to elicit indifferences. In the first experiment we, ultimately, asked people to state their indifference value. Such a procedure, in which people are asked directly to state their indifference value, is referred to as a matching task. In the second experiment, we asked subjects to make a series of choices and their indifference value was inferred from these choices. Several studies have shown that different elicitation procedures induce different cognitive processes and, consequently, can lead to different results (Tversky *et al.*, 1988, Bostic *et al.*, 1990, Fischer and Hawkins, 1993, Delquié, 1997). Tversky et al. (1988) have argued that preferences tend to be more lexicographic in choice behavior: people tend to focus on the most important attribute when making choices. In a matching task, people are more willing to make a trade-off between the attributes.

The subjects in the second experiment were 32 economics students at the University Pompeu Fabra, who were paid €36 for their participation. No student had participated in the first experiment. Responses were elicited in two personal interview sessions separated by two weeks. Prior to the actual experiment, the experimental design was tested in several pilot sessions using university staff as subjects.

The experimental procedure was similar to the procedure used in the first experiment except for the following. For each subject, we elicited two sequences $\{t_{1,q},\ldots,t_{6,q}\}$, one for years with back pain and one for years with migraine. We selected back pain and migraine because these are common illnesses and subjects were likely to know people suffering from these. The two sequences were elicited in different sessions to avoid that people would recall their earlier answers. The order in which the sequences were elicited varied across subjects.

We described the health states by the Maastricht Utility Measurement Questionnaire, a widely used instrument to describe health states in medical research (Rutten-van Mölken *et al.*, 1995). The description of the health states is available in Appendix F. In the migraine questions, subjects were told that on average they spend 5 days per month with migraine. The health descriptions were printed on cards, which were handed to the subjects.

As mentioned above, preferences were elicited by a sequence of choices. The wording of the questions was similar to the first experiment (see Appendix E), except, of course, that people were now told that the years were spent with back pain or migraine. For all subjects we started with a choice between $(55,t_{j-1,q})$ and $(45,\ t_{j-1,q})$, followed by a choice between $(55,t_{j-1,q})$ and $(45,\ t_{j-1,q} + 10)$. The stimuli in the subsequent choice questions depended on the answers to previous choice questions. After indifference was established, we displayed the final preference comparison again and we asked subjects to confirm indifference. If a subject did not confirm indifference, the elicitation was started anew.

Results on Constant Marginal Utility

For all subjects $t_{6,migraine}$ and $t_{6,back\ pain}$ were less than 45 years and, therefore, the analysis of Section 4 is valid for all subjects. Figure 2 shows the utility functions for life duration with back pain and with migraine based on median responses. Both utility functions appear to be concave in deviation from constant marginal utility. The null hypothesis of equal step sizes was rejected for both health states, both by analysis of variance ($p < 0.001$) and by the Friedman test ($p < 0.001$). The conclusions are the same if we use mean values instead of median values. These findings confirm the conclusion of the first experiment that constant marginal utility is violated and that the linear QALY model does not hold at the aggregate level.

{insert Figure 2 here}

Table 3 shows the results of the individual analyses. The procedure of classifying individuals is similar to the first experiment. The table shows that, for both health states, a clear majority of subjects violate constant marginal utility and have concave utility for life duration.

Table 3: Classification of Subjects According to the Shape of Their Utility Function

	Shape		
	Concave	Linear	Convex
Back Pain	75.0%	21.9%	0%
Migraine	78.1%	15.6%	0%

Results on Utility Independence

Utility independence could be tested within subjects by comparing the successive step sizes of the sequence $\{t_{1,\text{back pain}},\ldots,t_{6,\text{ back pain}}\}$ for years with back pain with those of the sequence $\{t_{1,\text{migraine}}\ldots,t_{6,\text{ migraine}}\}$ for years with migraine. The first step size differs significantly between the two sequences both by the paired t-test ($p = 0.030$) and by the nonparametric Wilcoxon signed-ranks test ($p = 0.032$). The other five step sizes do not differ significantly ($p > 0.10$ in all comparisons). Because we conducted 6 statistical tests using the same data set, the probability of falsely rejecting the null hypothesis of no difference is rather high. We, therefore, corrected for multiple significance testing both by the Bonferroni method and by Tukey's method for multiple comparisons. After correction for multiple significance testing, none of the differences is significant.

We obtain between-subjects tests of utility independence by comparing successive step sizes of the sequence for years in good health (the data from the first experiment) with those of the sequence for years with back pain and with those of the sequence for years with migraine. We cannot reject equality of the step sizes of the sequence for years in good health and those of the sequence for years with migraine ($p > 0.10$ in all comparisons). The first step size in the sequence for years in good health differs significantly from the first step size in the sequence for years with back pain both by the independent-samples t-test ($p = 0.039$) and by the nonparametric Mann-Whitney test ($p = 0.034$). The other step sizes do not differ significantly ($p > 0.10$ in all comparisons). After correction for multiple significance testing, none of the differences is significant.

Curve fitting

In the previous sections, we made no assumptions about the utility function for life duration. In this subsection, we analyze the data assuming specific parametric forms for the utility function for life duration. Parametric fitting has the advantage that irregularities in the data are smoothened out. A disadvantage is that the results may depend on the specific family chosen.

We examined three parametric forms, the power family, the exponential family, and the expo-power family. Let $z = x/t_{6,q}$, $x \in [0,t_{6,q}]$. The *power family* is defined by z^r if $r > 0$, by $\ln(z)$ if $r = 0$, and by $-z^r$ if $r < 0$. We only considered the case $r > 0$. The functions $\ln(z)$ and $-z^r$, $r < 0$, go to minus infinity if z goes to zero, implying that an individual is not prepared to run any risk of death, contrary to empirical observation. The *exponential family* is defined by $(e^{rz}-1)/(e^r-1)$ if $r > 0$, by z if $r = 0$, and by $(1-e^{rz})/(1-e^r)$ if $r < 0$. The power and exponential family are widely used in economics and (medical) decision analysis. The exponential family corresponds to the common procedure of discounting QALYs at a constant rate.

The expo-power family was introduced by Abdellaoui *et al.*, 2002) and is a variation of a two-parameter family proposed by Saha, 1993). The *expo-power family* is defined by $(1-\exp(-\frac{z^r}{r}))/(1-\exp(-1/r))$ with $r > 0$. We only considered the case $r > 0$, because the functions corresponding to $r = 0$ and to $r < 0$ go to minus infinity if z goes to zero. An important advantage of the expo-power family is that for $r \leq 1$ the function is concave and has both decreasing absolute risk aversion and increasing proportional risk aversion. These features are considered desirable in the economics literature and are supported by empirical evidence (Arrow, 1971, Binswanger, 1980, 1981, Rabin, 2000, Holt and Laury, 2002). Neither the power family nor the exponential family has both of these features.

Table 4: Parameter Estimates

	Parametric Families								
	Power			Exponential			Expo-Power		
	Median	Mean	St.Dev.	Median	Mean	St.Dev.	Median	Mean	St.Dev.
Good H.	0.72	0.74	0.15	−1.05	−1.03	0.60	0.99	1.00	0.16
BackPain	0.77	0.81	0.18	−0.83	−0.77	0.67	1.03	1.07	0.19
Migraine	0.73	0.75	0.14	−0.97	−1.01	0.67	0.99	1.01	0.15

For each individual we estimated the coefficients of the power, the exponential and the expo-power function by minimizing the sum of squared residuals. Table 4 shows the results. The parametric fittings reject linearity of the utility function ($p < 0.001$ in all tests), providing further evidence against the time-linear QALY model. The estimated parameters are rather different from those corresponding to the time-linear QALY model, suggesting that falsely assuming linear uility for life duration may lead to the wrong policy recommendations.

If utility independence holds then the parameters of the utility functions should be independent of health status. Table 4 shows that the parameters for good health and for migraine are close. The parameters for back pain are somewhat different. None of the differences between the parameters is, however, significant at the 5% level. No significant difference in goodness of fit could be detected between the three families.

6. Conclusion

In this paper, we performed new tests, that are robust to violations of expected utility, of two QALY models. Our findings reject the assumption that the utility for life duration is linear, both at the aggregate level and for a majority of subjects, but support the assumption that life duration is utility independent from health status, and hence, the time-nonlinear QALY model. In comparison with the time-linear QALY model, the time-nonlinear QALY model requires not only the elicitation of the health utility function but also the elicitation of the utility function for duration. As we show in this paper, this elicitation is feasible and the time-nonlinear QALY model, therefore, remains tractable for practical applications. Parametric estimations suggest that the utility function for duration is concave. The very finding of concave utility for duration is not surprising in itself. The contribution of the paper is to be the first to demonstrate this empirically without being invalidated by violations of expected utility.

Plausible explanations for concave utility of life duration are decreasing marginal utility, with people valuing additional life-years less the higher their life-expectancy, and the discounting of future utility. The good performance of the exponential family in our parametric fittings suggests that a QALY model with a constant rate of discount may describe people's preferences for health well. The validity of such a simple model would be useful for applications. Concavity of the utility for life duration may, however, also have arisen because, in spite of our instructions that health status was fixed, the subjects anticipated that quality of life would be lower at older ages. In this case our finding of concave utility for life duration could be an artifact.

Economic evaluation of health care is primarily a prescriptive exercise and expected utility is still the dominant prescriptive theory of decision under uncertainty (Kahneman and Tversky, 1979, p.277, Hammond, 1988). Even if utilities are to be used in a prescriptive analysis, their measurement, as it is commonly performed today, is still a descriptive exercise. It is, therefore, vulnerable to the biases induced by violations of expected utility. This paper shows how these biases can be avoided in tests of two important QALY models. Our findings suggest using the time-nonlinear QALY model in economic evaluations of health care.

Appendix A: Characterization of Expression (1) When the Outcome Set is Connected

Wakker (1993) characterized the additive representation $f \mapsto U_r(f_r) + U_s(f_s)$ for the case where the outcome set $\mathcal{H}$ is a connected topological space. If $\mathcal{H}$ is a topological space and $\mathcal{H}^2_{\downarrow}$ is endowed with the restriction of the product topology on $\mathcal{H}^2$ then $\succcurlyeq$ is *continuous* on $\mathcal{H}^2_{\downarrow}$ if for all $y \in \mathcal{H}$ the sets $\{x \in \mathcal{H}^2_{\downarrow}: x \succcurlyeq y\}$ and $\{x \in \mathcal{H}^2_{\downarrow}: x \preccurlyeq y\}$ are closed. The preference relation satisfies *outcome monotonicity* if for all acts f,g in $\mathcal{H}^2_{\downarrow}$ if $f_r \succcurlyeq g_r$ and $f_s \succcurlyeq g_s$ then $f \succcurlyeq g$, where the consequent preference is strict if either antecedent preference is strict. The *Thomsen condition* is satisfied on $\mathcal{H}^2_{\downarrow}$ if $(f_r, f_s) \sim (g_r, g_s)$ & $(h_r, g_s) \sim (f_r, h_t) \Rightarrow (h_r, f_s) \sim (g_r, h_t)$ whenever all six acts are contained in $\mathcal{H}^2_{\downarrow}$.

A health state $x \in \mathcal{H}$ is *maximal* if for no other health state $y \in \mathcal{H}$, $y \succ x$. A health state $x \in \mathcal{H}$ is *minimal* if for no other health state $y \in \mathcal{H}$, $x \succ y$. An *extreme act* either assigns to both states of nature r and s a maximal health state or to both states of nature r and s a minimal health state.

Lemma A.1

Let $\mathcal{H}$ be a connected topological space, and let $\succcurlyeq$ be a weak order on $\mathcal{H}^2_{\downarrow}$ that satisfies continuity, outcome monotonicity, and the Thomsen condition. Then there exist continuous functions U_r and U_s from $\mathcal{H}$ to $\mathbb{R}$ such that $f \mapsto U_r(f_r) + U_s(f_s)$ represents $\succcurlyeq$ on $\mathcal{H}^2_{\downarrow}\backslash\{$extreme acts$\}$. If U_r and U_s are linear with respect to each other, then the representation can be extended by continuity to the entire set $\mathcal{H}^2_{\downarrow}$. The functions U_r and U_s are unique up to a positive linear transformation with common units.

□

Proof

See Wakker (1993, Theorem 3.3(a), Proposition 3.5, and Remark 3.7). □

Appendix B: Extension to Outcome Sets that Are Not Connected

The set $\mathcal{Q}$ is general. Therefore, no topologies are naturally given on $\mathcal{H}$ and $\mathcal{H}^2$ and Lemma A.1 no longer applies. However, Lemma A.1 can be extended to a domain that is not connected, provided that the maximal connected subspaces of the domain, the topological components, overlap sufficiently in the preference order. The zero-condition, which was defined in Section 3, ensures this sufficient overlap.

If $\mathcal{H}$ is not a topological space, Wakker's definition of continuity is ambiguous. Instead, we assume that $\succcurlyeq$ is continuous in duration. For j = r,s, let $x_j f$ denote the prospect f with f_j replaced by x. The preference relation is *continuous in duration* if for all $f,g \in \mathcal{H}^2_\downarrow$, for $q \in \mathcal{Q}$ and for j = r,s, the sets $\{t \in \mathcal{T}: (q,t)_j f \succcurlyeq g\}$ and $\{t \in \mathcal{T}: (q,t)_j f \preccurlyeq g\}$ are closed.

The set of durations $\mathcal{T}$ is an interval, and hence the Euclidean topology is defined on $\mathcal{T}$. It is well known that the Euclidean topology is connected.

Lemma B.1:

Let $\succcurlyeq$ be a weak order on $\mathcal{H}^2_\downarrow$ that satisfies the zero condition, monotonicity in duration, continuity in duration, outcome monotonicity, and the Thomsen condition on $\mathcal{H}^2_\downarrow$. Then there exist functions U_r and U_s from $\mathcal{H}$ to $\mathbb{R}$ such that $f \mapsto U_r(f_r) + U_s(f_s)$ represents $\succcurlyeq$ on $\mathcal{H}^2_\downarrow$ \{extreme acts}. U_r and U_s are strictly increasing in duration and continuous in duration. If U_r and U_s are linear with respect to each other, then the representation can be extended by continuity to the entire set $\mathcal{H}^2_{\downarrow}$. The functions U_r and U_s are unique up to a positive linear transformation with common units.

□

Proof

Consider the order topology $T_{\succcurlyeq}$ on $\mathcal{H}$, i.e., the smallest topology containing all sets $\{h\in\mathcal{H}: h\succ g\}$ and $\{h\in\mathcal{H}: h\prec g\}$. The preference relation $\succcurlyeq$ on $\mathcal{H}$ is continuous with respect to this topology. By Lemma 3.1 in Bleichrodt and Miyamoto (2003), $T_{\succcurlyeq}$ is connected if the zero-condition holds. Then the product topology $T^2_{\succcurlyeq}$ on the set of acts $\mathcal{H}^2$ is also connected. By Lemma 3.2 in Bleichrodt and Miyamoto (2003), $\succcurlyeq$ on $\mathcal{H}^2_{\downarrow}$ is continuous with respect to $T^2_{\succcurlyeq}$. The preference relation $\succcurlyeq$ satisfies outcome monotonicity and the Thomsen condition. Hence, by Theorem 3.3 in Wakker (1993) there exist functions U_r and U_s from $\mathcal{H}$ to $\mathbb{R}$ such that $f \mapsto U_r(f_r) + U_s(f_s)$ represents $\succcurlyeq$ on $\mathcal{H}^2_{\downarrow}\backslash\{\text{extreme acts}\}$. U_r and U_s are continuous in duration by continuity in duration and strictly increasing in duration by monotonicity in duration. By proposition 3.5 in Wakker (1993), if U_r and U_s are linear with respect to each other, then the representation can be extended to the entire set $\mathcal{H}^2_{\downarrow}$ by continuity of $\succcurlyeq$ with respect to $T^2_{\succcurlyeq}$. By Theorem 3.3 in Wakker (1993), the functions U_r and U_s are unique up to a positive linear transformations with common units. □

Appendix C: Generalization to More Than Two States of Nature

Throughout the paper, we have assumed that there are only two states of nature. We now generalize our results to the case where the number of states of nature is arbitrary, but finite. Let $\mathcal{S} = \{1,\ldots,m\}$ be the state space. We assume that *outcome monotonicity* holds in the sense that if $f_j \succcurlyeq g_j$ for all $j \in \mathcal{S}$ then $f \succcurlyeq g$ with $f \succ g$ if at least one antecedent preference is strict. Hence, there are no null states. We further assume that there exist additive functions $V_j:\mathcal{H}\to\mathbb{R}$, $j\in\mathcal{S}$, such that $f \mapsto \sum_{j\in\mathcal{S}} V_j(f_j)$ represents $\succcurlyeq$ on $\mathcal{H}_{\downarrow}^{n}$. For any proper subset $A \subset \mathcal{S}$, let (x, A, y) denote the act that gives outcome x if $s_j \in A$ and y otherwise. Let $\mathcal{H}_{\downarrow}^{A}$ be the set of all such acts with $x \succcurlyeq y$. $\mathcal{H}_{\downarrow}^{A}$ is isomorphic to $\mathcal{H}_{\downarrow}^{2}$ under the map $\varphi((x, A, y)) = (x,y)$. Hence, the conditions identified in Theorems 3.1 and 3.2 can be applied to $\mathcal{H}_{\downarrow}^{A}$ to give the time-linear and the time-nonlinear QALY model, respectively. The implication that the QALY models imply the preference conditions on each $\mathcal{H}_{\downarrow}^{A}$, and in fact on the entire domain, are easy to verify and are left to the reader.

Appendix D: Proofs

Proof of Theorem 3.1

It is easily verified that (ii) implies (i). Suppose that (i) holds. Lemma B.1 ensures that Expression (1) holds. Hence, we can apply the proof of Theorem 2 in Bleichrodt and Quiggin (1997) to derive (ii). □

Proof of Theorem 3.2.

Suppose that (ii) holds. Because V(q) is everywhere positive, it is easily verified that duration is utility independent of health status. The verification of the other conditions is straightforward.

Suppose that (i) holds. By Lemma B.1 and continuity of $\succcurlyeq$ with respect to $T_{\succcurlyeq}^{2}$, the theorem holds in general if it holds in case $\mathcal{H}$ contains no maximal or minimal outcomes.

Therefore, assume that $\mathcal{H}$ contains no maximal or minimal outcomes. Define $\mathcal{T}^2_{\downarrow} = \{(t_1,t_2) \in\mathcal{T}^2 : t_1 \geq t_2\}$. By monotonicity in duration, $(t_1,t_2)\in \mathcal{T}^2_{\downarrow} \Leftrightarrow ((q,t_1),(q,t_2))\in \mathcal{H}^2_{\downarrow}$. Define the relation $\succcurlyeq_t$ on $\mathcal{T}^2_{\downarrow}$ by $(t_1,t_2) \succcurlyeq_t (t_3,t_4)$ if for some $q\in\mathcal{Q}$ $((q,t_1),(q,t_2)) \succcurlyeq ((q,t_1),(q,t_2))$. Because duration is utility independent the choice of q is immaterial. Choose an arbitrary $q'\in\mathcal{Q}$ and define functions W_r and W_s from $\mathcal{T}$ to $\mathbb{R}$ by $W_r(t_1) = U_r(q',t_1)$ and $W_s(t_1) = U_s(q',t_1)$. Duration being utility independent on $\mathcal{H}^2_{\downarrow}$ implies that for any $q\in\mathcal{Q}$ both the function $(t_1,t_2)\mapsto U_r(q,t_1) + U_s(q,t_2)$ and the function $(t_1,t_2)\mapsto W_r(t_1) + W_s(t_2)$ represent $\succcurlyeq_t$ on $\mathcal{T}^2_{\downarrow}$. By the uniqueness properties of U_r and U_s, there exist real q_r, q_s and positive $V(q)$ such that for all $t\in\mathcal{T}$:

$$U_r(q,t) = V(q)\cdot W_r(t) + q_r \tag{D.1}$$

$$U_s(q,t) = V(q)\cdot W_s(t) + q_s \tag{D.2}$$

Set $W_r(0) = W_s(0) = 0$, which is allowed by the uniqueness properties of U_r and U_s and the zero-condition. Now, $0 = W_r(0) = W_s(0) = U_r(q',0) = U_s(q',0) = U_r(q,0) = U_s(q,0)$ where the latter two equalities follow by the zero-condition. Therefore, $V(q)\cdot 0 + q_r = 0$ from which it follows that for all $q\in\mathcal{Q}$, $q_r = 0$. Similarly it can be shown that for all $q\in\mathcal{Q}$, $q_s = 0$. By outcome monotonicity, $U_r(q,t_1) = V(q)\cdot W_r(t_1)$ and $U_s(q,t_1) = V(q)\cdot W_s(t_1)$ represent the same preference relation on $\mathcal{H} = \mathcal{Q}\times\mathcal{T}$. Hence, by the uniqueness properties of multiplicative representations (Krantz *et al.*, 1971) there exist λ, $\gamma > 0$ such that for all $(q,t)\in\mathcal{H}$, $U_r(q,t) = V(q)\cdot W_r(t) = \lambda\cdot V(q)^\gamma\cdot W_s(t)^\gamma = \lambda\cdot U_s(q,t)^\gamma$.

Because not all health states are equivalent, $V(q)$ is not constant. Let $q_1,q_2\in\mathcal{Q}$ be such that $V(q_1) \neq V(q_2)$. Then for arbitrary $t\in\mathcal{T}$, $t\neq 0$, $V(q_1)\cdot W_r(t) = \lambda\cdot V(q_1)^\gamma\cdot W_s(t)^\gamma$ and $V(q_2)\cdot W_r(t) = \lambda\cdot V(q_2)^\gamma\cdot W_s(t)^\gamma$. Substitution gives $\frac{V(q_1)}{V(q_2)} = \left(\frac{V(q_1)}{V(q_2)}\right)^\gamma$. Hence, $\gamma = 1$ and U_r and U_s are linear with respect to each other. Define $W(t) = W_s(t)$, $U(q,t) = V(q)\cdot W(t)$, $\pi_r = \frac{\lambda}{\lambda+1}$ and $\pi_s = \frac{1}{\lambda+1}$.

Then the nonlinear QALY model holds and $\pi_r U$ and $\pi_s U$ are additive utility functions for $\succcurlyeq$ on $\mathcal{H}^2_{\downarrow}$. □

Appendix E: Question Wording in the First Experiment.

Suppose that you have been diagnosed to have symptoms of one of two diseases: A or B. From medical experience it is known that half of the people with these symptoms have disease A and half have disease B. There exist two treatments for these diseases, but the effects of the treatments depend on which disease you have. To be effective, the treatments have to start immediately. Unfortunately, it is only known which disease you have after treatment has started. That is, you have to choose which treatment to undergo when you are still uncertain which disease you have.

In the following questions you are faced with different outcomes of the treatments. In each question you are asked to state the number of life years for which you consider the two treatments equivalent. Suppose in every question that you spend the years in good health. The way of presentation is as follows:

Treatment	Disease A	Disease B
1	50	5
2	25	x_i

Suppose you choose to undergo treatment 1, then you live for 50 more years if you turn out to have disease A. If, on the other hand, you turn out to have disease B you live for 5 more years. If you choose to undergo treatment 2 you live for 25 more years if you turn out to have disease A. If, on the other hand, you turn out to have disease B you live for x_i more years. In the following questions you will be asked to indicate the number of years x_i for which you consider the two treatments equivalent. In the above example, it is plausible that x_i lies between 5 years and 50 years. If $x_i = 5$ years then treatment 1 is clearly better than treatment 2. If $x_i = 50$ years then treatment 2 is clearly better than treatment 1.

Now consider the following question:

Question 1.

If you choose treatment 1 and you turn out to have disease A you live for 55 more years in good health, but if you turn out to have disease B you die immediately. If you choose treatment 2 and you turn out to have disease A you live for 45 more years in good health, but if you turn out to have disease B you live for x_1 more years in good health. Choose the value of x_1 for which you consider the two treatments equivalent and put this value on your answer sheet.

Treatment	Disease A	Disease B
1	55	0
2	45	x_1

Appendix F: The Description of the Health States in the Second Experiment

Back Pain

Unable to perform some tasks at home and/or at work

Able to perform all self care activities (eating, bathing, dressing) albeit with some difficulties

Unable to participate in many types of leisure activities

Often moderate to severe pain and/or other complaints

Migraine

Unable to perform usual tasks at home and/or at work

Able to perform all self care activities (eating, bathing, dressing) albeit with some difficulties

Unable to participate in any type of leisure activity

Severe headache

Acknowledgements

Antonio Cabrales, David de Meza (the editor), Enrico Diecidue, Eddy van Doorslaer, Erik Schut, Anne Spencer, Peter Wakker, and two anonymous referees gave many helpful comments. Han Bleichrodt's research was made possible by a fellowship from the Royal Netherlands Academy of Arts and Sciences and by a grant from the Netherlands Organisation for Scientific Research (NWO).

References:

Abdellaoui, M., Barrios, C., and Wakker, P. P. (2002). 'Reconciling introspective utility with revealed preference: Experimental arguments based on prospect theory', vol. pp. Working Paper, ENS, Cachan.

Arrow, K. J. (1971). *Essays in the theory of risk-bearing*. Amsterdam: North Holland.

Binswanger, H. P. (1980). 'Attitude toward risk: Experimental measurement in rural india', *American Journal of Agricultural Economics,* vol. **62**, pp. 395-407.

Binswanger, H. P. (1981). 'Attitudes towards risk: Theoretical implications of an experiment in rural india', *The Economic Journal,* vol. **91**, pp. 867-890.

Bleichrodt, H. and Miyamoto, J. (2003). 'A characterization of quality-adjusted life-years under cumulative prospect theory', *Mathematics of Operations Research,* vol. **28**, pp. 181-193.

Bleichrodt, H. and Quiggin, J. (1997). 'Characterizing QALYs under a general rank dependent utility model', *Journal of Risk and Uncertainty,* vol. **15**, pp. 151-165.

Bleichrodt, H., Wakker, P. P., and Johannesson, M. (1997). 'Characterizing QALYs by risk neutrality', *Journal of Risk and Uncertainty,* vol. **15**, pp. 107-114.

Bostic, R., Herrnstein, R. J., and Luce, R. D. (1990). 'The effect on the preference reversal of using choice indifferences', *Journal of Economic Behavior and Organization,* vol. **13**, pp. 193-212.

Camerer, C. (1995). 'Individual decision making', in: (J. Kagel andA. Roth, eds.), *The handbook of experimental economics*, Princeton, NJ: Princeton University Press.

Crawford, V. P. (1990). 'Equilibrium without independence', *Journal of Economic Theory,* vol. **50**, pp. 127-154.

Cubitt, R. P. and Sugden, R. (1998). 'The selection of preferences through imitation', *Review of Economic Studies,* vol. **65**, pp. 761-771.

de Wit, G. A., van Busschbach, J. J., and de Charro, F. T. (2000). 'Sensitivity and perspective in the valuation of health status', *Health Economics,* vol. **9**, pp. 109-126.

Dekel, E., Safra, Z., and Segal, U. (1991). 'Existence and dynamic consistency of Nash equilibrium with non-expected utility preferences', *Journal of Economic Theory,* vol. **55**, pp. 302-318.

Delquié, P. (1997). 'Bi-matching': A new preference assessment method to reduce compatibility effects', *Management Science,* vol. **43**, pp. 640-658.

Fischer, G. W. and Hawkins, S. A. (1993). 'Strategy compatibility, scale compatibility, and the prominence effect', *Journal of Experimental Psychology: Human Perception and Performance,* vol. **19**, pp. 580-597.

Gul, F. (1991). 'A theory of disappointment aversion', *Econometrica,* vol. **59**, pp. 667-686.

Hammond, P. J. (1988). 'Consequentialist foundations for expected utility', *Theory and Decision,* vol. **25**, pp. 25-78.

Hertwig, R. and Ortmann, A. (2001). 'Experimental practices in economics: A methodological challenge for psychologists?' *Behavioral and Brain Sciences,* vol. **24**, pp. 383-451.

Holt, C. A. and Laury, S. K. (2002). 'Risk aversion and incentive effects', *American Economic Review,* vol. **92**, pp. 1644-1655.

Kahneman, D. and Tversky, A. (1979). 'Prospect theory: An analysis of decision under risk', *Econometrica,* vol. **47**, pp. 263-291.

Karni, E. (1985). *Decision-making under uncertainty: The case of state-dependent preferences*. Cambridge, Massachusetts: Harvard University Press.

Karni, E. and Safra, Z. (1989). 'Dynamic consistency, revelations in auctions and the structure of preferences', *Review of Economic Studies,* vol. **56**, pp. 421-434.

Krantz, D. H., Luce, R. D., Suppes, P., and Tversky, A. (1971). *Foundations of measurement, vol. 1*. New York: Academic Press.

Machina, M. J. (1995). 'Non-expected utility and the robustness of the classical insurance paradigm', *Geneva Papers on Risk and Insurance Theory,* vol. **20**, pp. 9-50.

McNeil, B. J., Weichselbaum, R., and Pauker, S. G. (1978). 'Fallacy of the five-year survival in lung cancer', *New England Journal of Medicine,* vol. **299**, pp. 1397-1401.

McNeil, B. J., Weichselbaum, R., and Pauker, S. G. (1981). 'Tradeoffs between quality and quantity of life in laryngeal cancer', *New England Journal of Medicine,* vol. **305**, pp. 982-987.

Miyamoto, J. M. and Eraker, S. A. (1988). 'A multiplicative model of the utility of survival duration and health quality', *Journal of Experimental Psychology: General,* vol. **117**, pp. 3-20.

Miyamoto, J. M. and Wakker, P. P. (1996). 'Multiattribute utility theory without expected utility foundations', *Operations Research,* vol. **44**, pp. 313-326.

Miyamoto, J. M., Wakker, P. P., Bleichrodt, H., and Peters, H. J. M. (1998). 'The zero-condition: A simplifying assumption in qaly measurement and multiattribute utility', *Management Science,* vol. **44**, pp. 839-849.

Quiggin, J. (1981). 'Risk perception and risk aversion among Australian farmers', *Australian Journal of Agricultural Economics,* vol. **25**, pp. 160-169.

Rabin, M. (2000). 'Risk aversion and expected-utility theory: A calibration theorem', *Econometrica,* vol. **68**, pp. 1281-1292.

Rutten-van Mölken, M. P., Bakker, C. H., van Doorslaer, E. K. A., and van der Linden, S. (1995). 'Methodological issues of patient utility measurement. Experience from two clinical trials', *Medical Care,* vol. **33**, pp. 922-937.

Saha, A. (1993). 'Expo-power utility: A 'flexible' form for absolute and relative risk aversion', *American Journal of Agricultural Economics,* vol. **75**, pp. 905-913.

Schmeidler, D. (1989). 'Subjective probability and expected utility without additivity', *Econometrica,* vol. **57**, pp. 571-587.

Starmer, C. (2000). 'Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk', *Journal of Economic Literature,* vol. **28**, pp. 332-382.

Stiggelbout, A. M., Kiebert, G. M., Kievit, J., Leer, J. W. H., Stoter, G., and de Haes, J. C. J. M. (1994). 'Utility assessment in cancer patients: Adjustment of time tradeoff scores for the utility of life years and comparison with standard gamble scores', *Medical Decision Making,* vol. **14**, pp. 82-90.

Tversky, A. and Kahneman, D. (1992). 'Advances in prospect theory: Cumulative representation of uncertainty', *Journal of Risk and Uncertainty,* vol. **5**, pp. 297-323.

Tversky, A., Sattath, S., and Slovic, P. (1988). 'Contingent weighting in judgment and choice', *Psychological Review,* vol. **95**, pp. 371-384.

Verhoef, L. C. G., de Haan, A. F. J., and van Daal, W. A. J. (1994). 'Risk attitude in gambles with years of life: Empirical support for prospect theory', *Medical Decision Making,* vol. **14**, pp. 194-200.

Wakker, P. P. (1993). 'Additive representations on rank-ordered sets II. The topological approach', *Journal of Mathematical Economics,* vol. **22**, pp. 1-26.

Wakker, P. P. and Deneffe, D. (1996). 'Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown', *Management Science,* vol. **42**, pp. 1131-1150.

Wakker, P. P., Thaler, R. H., and Tversky, A. (1997). 'Probabilistic insurance', *Journal of Risk and Uncertainty,* vol. **15**, pp. 7-28.

Yaari, M. E. (1987). 'The dual theory of choice under risk', *Econometrica,* vol. **55**, pp. 95-115.

Figure 1: The Utility Function for Years in Good Health

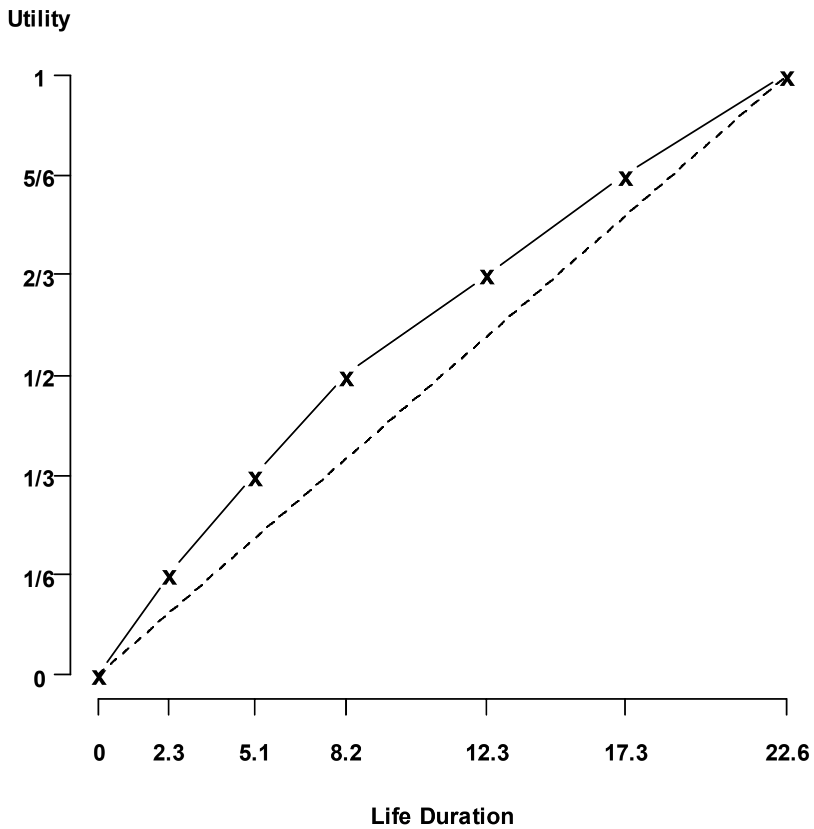

Figure 2: Utility Functions Elicited in the Second Experiment