

**Resolving Inconsistencies in Utility Measurement under Risk:
Tests of Generalizations of Expected Utility**

Han Bleichrodt, Erasmus University Rotterdam, The Netherlands.

Jose Maria Abellan-Perpiñan, University of Murcia, Spain.

Jose Luis Pinto, University of Pompeu Fabra, Barcelona, Spain.

Ildefonso Mendez-Martinez, University of Murcia, Spain.

January 2005

Abstract

This paper explores biases in the elicitation of utilities under risk and the contribution that generalizations of expected utility can make to the resolution of these biases. We used five methods to measure utilities under risk and found clear violations of expected utility. Of the theories studied, prospect theory was most consistent with our data. The main improvement of prospect theory over expected utility was in comparisons between a riskless and a risky prospect (riskless-risk methods). We observed no improvement over expected utility in comparisons between two risky prospects (risk-risk methods). An explanation why we found no improvement of prospect theory over expected utility in risk-risk methods may be that there was less overweighting of small probabilities in our study than has commonly been observed.

KEYWORDS: Utility Measurement, Nonexpected Utility, Prospect Theory, Health.

Address correspondence to: Han Bleichrodt, iMTA/iBMG, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. Email: bleichrodt@bmg.eur.nl.

Acknowledgements: We are grateful to Peter Wakker for his helpful comments on a previous draft and to George Wu for sending us additional statistics on the data in Wu and Gonzalez (1996). Han Bleichrodt's research was made possible by a VIDI-grant from the Netherlands Organization for Scientific Research (NWO). José Luis Pinto acknowledges funding from project BBVA1-04X.

1. Introduction

This paper explores biases in the elicitation of utilities under risk and the contribution that generalizations of expected utility make to the resolution of these biases. Risky utilities are widely used in decision analysis. The main examples are the probability equivalence (PE) method, where people are asked to state a probability in a risky prospect that makes them indifferent between this risky prospect and a given outcome for sure, and the certainty equivalence (CE) method, where people are asked to state an outcome for sure that makes them indifferent between this outcome and a given risky prospect. The appealing feature of using methods involving risk is that they allow to incorporate people's attitudes towards risk into decision analysis.

The common way to analyze the responses to utility measurements under risk is by assuming expected utility. A rationale for adopting expected utility is that decision analysis is essentially a prescriptive exercise and that expected utility is the dominant prescriptive theory of decision under risk. The problem with this point of view is that in most practical applications measuring utilities is a descriptive task, and the descriptive deficiencies of expected utility are widely documented (Starmer 2000). Using expected utility to analyze responses to utility measurement tasks in spite of its poor descriptive record carries the danger that biased utilities result and that decision analyses based on these biased utilities make incorrect recommendations. In fact, several studies have shown that utility measurement based on expected utility leads to inconsistencies. Hershey and Schoemaker (1985) showed that PE measurements lead to systematically higher utilities than CE measurements. McCord and de Neufville (1986) found that the utility function that is elicited by the CE method depends on the value at which probability is kept fixed. In the health domain, several authors have shown that different procedures to elicit utilities by the PE method that according to expected utility are equivalent lead to systematically different results (Llewellyn-Thomas et al. 1982, Rutten-van Mólken et al. 1995, Bleichrodt 2001, Oliver 2003).

A solution for the observed inconsistencies was suggested by Bleichrodt et al. (2001). They argued that people's preferences are affected by biases. The biases they considered were probability weighting, the nonlinear evaluation of probabilities, and loss aversion, the finding that people are more sensitive to losses than to gains of the same size. Both biases have been modeled by prospect theory (Kahneman and Tversky 1979, Tversky and Kahneman 1992), currently the main descriptive theory of decision under risk. Bleichrodt et al. proposed new formulas based on prospect theory to evaluate answers to PE and CE measurements and showed that these new formulas were able to resolve the systematic discrepancy between PE and CE utilities.

It is not clear whether the findings of Bleichrodt et al. (2001) can be extended to other methods than the PE and the CE. The PE and CE are examples of methods in which a riskless prospect is compared with a risky prospect (*riskless-risk methods*). Several empirical studies have observed that violations of expected utility primarily occur when one of the prospects under consideration is riskless; expected utility's descriptive record is much better when both prospects are risky (Conlisk 1989, Camerer 1992, Harless and Camerer 1994, Wu and Gonzalez 1996, Starmer 2000). Hence, it is of interest to examine the performance of prospect theory when we include methods that compare two risky prospects (*risk-risk methods*). It would be important for practical decision analysis in case we observed that expected utility could explain the data from risk-risk measurements. Instead of having to elicit probability weighting and loss aversion parameters, which is time consuming, preferences could then be measured under expected utility by using lottery equivalence methods (McCord and de Neufville 1986).

The purpose of this paper is to present more extensive data on inconsistencies in measuring utilities under risk. We compared five elicitation methods, three riskless-risk methods and two risk-risk methods. We evaluated these five methods under expected utility and under several generalizations of expected utility. The generalizations we considered were prospect theory, rank-dependent utility (Quiggin 1981), disappointment aversion (Gul 1991), and two recently proposed gambling effect models (Bleichrodt and Schmidt 2002, Diecidue et al. 2004).

An interesting feature of the gambling effect models is that they predict that expected utility will lead to consistent measurements in risk-risk methods.

We studied the performance of these theories in the health domain. The main reason we focused on health is practical relevance. Health is an important area of applied decision analysis (the majority of applied decision analyses are in the health domain, see Keller and Kleinmuntz 1998 and Smith and von Winterfeldt 2004) and risky methods are widely used in medical decision analysis. Of the theories we examined, prospect theory was most consistent with the data. The main improvement of prospect theory over expected utility was in riskless-risk methods. For risk-risk methods we found no violations both for expected utility and for prospect theory and, hence, prospect theory did not improve over expected utility.

In what follows, Section 2 introduces notation and describes the nonexpected utility theories studied in this paper. Section 3 describes the five elicitation methods used. Section 4 analyzes the predicted deviations from expected utility and shows how utilities were computed under the theories considered. Section 5 describes the experiment that we performed and Section 6 its results. Section 7 discusses the main findings, implications, and limitations of our study and concludes the paper. Derivations are in the appendix.

2. Notation and outline of the models

We study preferences over chronic health states. Chronic health states will be written as (Q,T) , denoting T years in health state Q . Because the experiment reported later only invokes prospects with at most two different chronic health states, we will restrict the analysis to such *binary prospects*. Let $(p:(Q_1,T_1); (Q_2,T_2))$ denote the prospect that gives outcome (Q_1,T_1) with probability p and outcome (Q_2,T_2) with probability $1-p$. A prospect is *riskless* if $p = 1$; otherwise it is *risky*.

Preferences over prospects are denoted as usual: the relation \succsim denotes weak preference, \succ denotes strict preference, and \sim denotes indifference. Preferences over outcomes correspond

with preferences over riskless prospects. We assume throughout that prospects are *rank-ordered*, i.e. it is implicit in the notation $(p:(Q_1, T_1); (Q_2, T_2))$ that $(Q_1, T_1) \succcurlyeq (Q_2, T_2)$. *Expected utility* holds if there exists a function U from the set of chronic health states to the real numbers, called the *utility function*, such that prospects $(p:(Q_1, T_1); (Q_2, T_2))$ are evaluated by $pU(Q_1, T_1) + (1-p)U(Q_2, T_2)$ and preferences and choices correspond with this evaluation.

The gambling effect models of Bleichrodt and Schmidt (2002) and Diecidue et al. (2004) deviate from expected utility by assuming that there is not one utility function over outcomes but two. In the model of Diecidue et al. (2004) preferences are *prospect-dependent*: if a prospect is risky its outcomes are evaluated by a utility function U , if it is riskless its outcomes are evaluated by a utility function V . Diecidue et al. (2004) assume that $V(\cdot) = U(\cdot) + C(\cdot)$, where the function C denotes the benefit of certainty. To account for the commonly observed violations of expected utility, C must be positive for all outcomes.

In the model of Bleichrodt and Schmidt (2002) preferences are *context-dependent*: when both prospects in a comparison are risky its outcomes are evaluated by a utility function U , otherwise the outcomes of both prospects are evaluated by a utility function V . To account for the commonly observed violations of expected utility, U is assumed to be less concave than V . To illustrate the difference between the model of Diecidue et al. (2004) and that of Bleichrodt and Schmidt (2002) consider the indifference between the riskless prospect (Q_2, T_2) and the risky prospect $(p:(Q_1, T_1); (Q_2, T_2))$, $p \in (0, 1)$. According to the model of Diecidue et al. (2004) this indifference implies that

$$V(Q_2, T_2) = pU(Q_1, T_1) + (1-p)U(Q_3, T_3), \quad (1)$$

whereas according to the model of Bleichrodt and Schmidt (2002) this indifference implies that

$$V(Q_2, T_2) = pV(Q_1, T_1) + (1-p)V(Q_3, T_3). \quad (2)$$

2.1. Prospect theory, Rank-Dependent Utility, and Disappointment Aversion

Prospect theory deviates in three important respects from expected utility. First, carriers of value are gains and losses relative to a *reference point*. The location of the reference point is exogenously given and is not specified by prospect theory. We will denote the reference point as (Q_0, T_0) . In the formal analysis of Kahneman and Tversky (1979), where there is only one fixed reference point, the reference point is assigned utility zero. In this paper we will consider variations in the reference point and, hence, we do not follow this convention. A second deviation from expected utility is that people do not evaluate probabilities linearly, but weight probabilities; probability weighting for gains can be different from probability weighting for losses. Third, people are more sensitive to losses than to absolutely commensurate gains, a phenomenon known as *loss aversion*.

We will say that a prospect is *mixed* if it involves both a gain and a loss. A mixed prospect $((Q_1, T_1), p, (Q_2, T_2))$, $(Q_1, T_1) \succeq (Q_0, T_0) \succeq (Q_2, T_2)$, is evaluated as

$$\begin{aligned} PT((Q_1, T_1), p, (Q_2, T_2)) = & U(Q_0, T_0) + w^+(p)(U(Q_1, T_1) - U(Q_0, T_0)) \\ & - \lambda w^-(1-p)(U(Q_0, T_0) - U(Q_2, T_2)), \end{aligned} \quad (3)$$

where w^+ and w^- are the probability weighting functions for gains and losses and λ is a loss aversion parameter. The probability weighting functions assign 0 to 0 and 1 to 1 and are strictly increasing. We separate loss aversion from utility, because we will consider varying reference points and we want to establish a link with expected utility. Therefore, the utility function U reflects the intrinsic utility of chronic health states. Our method for modeling loss aversion is similar to Shalev (2000). In studies where the reference point is held fixed, loss aversion is often

incorporated in the utility function (Kahneman and Tversky 1979, Tversky and Kahneman 1992). In Eq.(3), outcomes are evaluated as deviations from the reference point through terms $U(Q_1, T_1) - U(Q_0, T_0)$ so as to combine the psychology of prospect theory with the utility function U of expected utility.

If both outcomes are gains, $(Q_1, T_1) \succcurlyeq (Q_2, T_2) \succcurlyeq (Q_0, T_0)$, then

$$\begin{aligned} PT((Q_1, T_1), p, (Q_2, T_2)) = & U(Q_0, T_0) + w^+(p)(U(Q_1, T_1) - U(Q_0, T_0)) \\ & + (1-w^+(p))(U(Q_2, T_2) - U(Q_0, T_0)). \end{aligned} \quad (4)$$

In case all outcomes are gains, prospect theory is identical to rank-dependent utility (Quiggin 1981). When $w^+(p) = \frac{p}{1+(1-p)\delta}$, Eq.(4) is equivalent to Gul's (1991) theory of disappointment aversion. The parameter $\delta \in (-1, \infty)$ reflects disappointment aversion.

If both outcomes are losses $(Q_0, T_0) \succcurlyeq (Q_1, T_1) \succcurlyeq (Q_2, T_2)$, then

$$\begin{aligned} PT((Q_1, T_1), p, (Q_2, T_2)) = & U(Q_0, T_0) - \lambda w^-(1-p)(U(Q_0, T_0) - U(Q_2, T_2)) \\ & - \lambda(1-w^-(1-p))(U(Q_0, T_0) - U(Q_1, T_1)). \end{aligned} \quad (5)$$

We assume throughout that changes in the reference point leave the probability weighting functions w^+ and w^- , the utility function U , and the loss aversion parameter λ unchanged.

Empirical studies have shown that the most common pattern for the probability weighting function is inverse S-shaped, overweighting small probabilities and underweighting intermediate and large probabilities (Tversky and Kahneman 1992, Tversky and Fox 1995, Gonzalez and Wu 1999, Abdellaoui 2000, Bleichrodt and Pinto 2000). Tversky and Kahneman (1992) proposed the following one-parameter functional form for the probability weighting function:

$$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}}, \quad (6)$$

which has an inverse S-shape for γ between 0.27 and 1. Tversky and Kahneman (1992) found a median value for γ^+ (the parameter for gains) of 0.61 and for γ^- (the parameter for losses) of 0.69. Later studies found comparable values for γ^+ and γ^- . These values imply that the probability for which the probability weighting functions change from overweighting probabilities to underweighting probabilities, i.e. the probability for which $w(p) = p$, lies between 0.30 and 0.40. Tversky and Kahneman (1992) also estimated the loss aversion coefficient and found a median value for λ equal to 2.25. Bleichrodt et al. (2001), who used the same definition of loss aversion as Tversky and Kahneman (1992) found a comparable value for λ .

3. Elicitation methods

Table 1 shows the elicitation methods that were used; the response that was elicited is printed in bold.

Table 1: The Five Elicitation Methods

Method	Question
Probability Equivalence (PE)	$(Q,T) \sim (\mathbf{p}':(FH,T); \text{Death})$
Certainty Equivalence (CE)	$(Q, \mathbf{T}_{ce}) \sim (\mathbf{p}':(FH,T); \text{Death})$
Value Equivalence (VE)	$(Q,T) \sim (\mathbf{p}':(FH, \mathbf{T}_{ve}); \text{Death})$
Probability Lottery Equivalence (PLE)	$(0.35:(Q,T); \text{Death}) \sim (\mathbf{r}':(FH,T); \text{Death})$
Value Lottery Equivalence (VLE)	$(0.35:(Q,T); \text{Death}) \sim (0.35:(FH, \mathbf{T}_{ve}); \text{Death})$

Note: FH stands for “full health”.

To enhance comparability between the riskless-risk methods, the selected value of the *gauge duration* T was the same in all five methods and the response p' elicited in the PE method was also used in the CE and the VE method. The use of probability 0.35, for which previous studies observed little probability distortion, in the risk-risk methods does not mean that probability weighting plays no role in these methods as r' would generally be lower than 0.35.

4. Predictions

We assumed throughout that people prefer more life duration to less, both in full health and in health state Q . Then expected utility, rank-dependent utility, disappointment aversion, and the gambling effect models predict that, except for random error, we should observe that $T = T_{ce} = T_{ve}$.¹ The two risk-risk methods could not be directly compared with each other and with the riskless-risk methods because they involved different probabilities.

The predictions made by prospect theory depend on the location of the reference point. Hershey and Schoemaker (1985) and Bleichrodt et al (2001) argued that in a utility elicitation task in which a subject compares two prospects and has to determine a parameter (either a probability or an outcome) that makes him indifferent between the two prospects, he will take an outcome of the prospect which parameters do not change as his reference point. Their argument is corroborated by empirical evidence (Stalmeier and Bezembinder 1999, Morrison 2000, Bleichrodt et al., 2001, Robinson et al. 2001). In the PE and in the VE this implies that the reference point is (Q, T) . In the CE the reference point is either (FH, T) or Death. The data in Bleichrodt et al. (2001) suggest that people took death as their reference point in the CE questions. Under the above hypotheses we would expect that $T = T_{ve}$. However, under prospect theory T_{ce} may well differ from T . Under the aforementioned hypotheses, the PE and VE method

¹ This follows from transitivity and the assumption that more life duration is preferred to less. Bleichrodt and Schmidt's (2002) model satisfies transitivity for riskless-risk methods, even though it does not satisfy transitivity in general.

compare mixed prospects, whereas the CE compares either two prospects involving only gains or two prospects involving only losses. It is well-known that people are more risk averse for mixed prospects and, hence, we expect $T_{ce} > T$. Again, because the risk-risk methods used different probabilities they do not yield specific predictions unless additional assumptions were made.

To be able to compute health state utilities we assumed in all models *multiplicativity* of U : $U(Q,T) = H(Q)L(T)$ where H and L are real-valued utility functions over the set of health states and the set of life durations, respectively. Throughout, we used the scaling $U(\text{Death}) = 0$ and $H(\text{FH}) = 1$.

Empirical support for multiplicativity was obtained by Miyamoto and Eraker (1988), Doctor et al. (2004), and Bleichrodt and Pinto (2004). A test of multiplicativity is obtained by comparing the answers to the PE questions for different gauge durations with each other and by comparing the answers to the PLE questions for different gauge durations with each other. Multiplicativity implies that, except for random response error, we should find the same probabilities p' in the different PE questions and the same probabilities r' in the different PLE questions.

To be able to compute $H(Q)$ from the responses to the CE, VE, and VLE questions, we had to assume a specific form for $L(T)$. We first assumed that L is linear in which case $U(Q,T)$ is equal to the QALY model, the most widely used model in medical decision analysis. The first row of Table 2 shows how $H(Q)$ was computed under expected utility with linear utility. The entries of Table 2 are derived in Appendix A. There is little empirical support for the assumption that utility is linear in life duration; an exception is the study by Doctor et al. (2004). We, therefore, subsequently assumed that L is a power function. The power function is often used in decision analysis and several studies have observed that it yields a good fit in the health domain (e.g. Pliskin et al. 1980, Miyamoto and Eraker 1985, Stiggelbout et al. 1994, Cher et al. 1997). The second row of Table 2 shows the effect on $H(Q)$ of replacing the assumption of linear utility

by the assumption of power utility. The table shows that the PE and the PLE method are not affected by the choice of L .

Table 2: Utilities under various assumptions

	PE	CE	VE	PLE	VLE
EU-linear	p'	$p' \frac{T}{T_{ce}}$	$p' \frac{T_{ve}}{T}$	$\frac{r'}{0.35}$	$\frac{T_{vle}}{T}$
EU-power	p'	$p' \left(\frac{T}{T_{ce}}\right)^\beta$	$p' \left(\frac{T_{ve}}{T}\right)^\beta$	$\frac{r'}{0.35}$	$\left(\frac{T_{vle}}{T}\right)^\beta$
DSW	$\frac{p'}{\alpha}$	$\frac{p'}{\alpha} \left(\frac{T}{T_{ce}}\right)^\beta$	$\frac{p'}{\alpha} \left(\frac{T_{ve}}{T}\right)^\beta$	$\frac{r'}{0.35}$	$\left(\frac{T_{vle}}{T}\right)^\beta$
BS	$p'^{1/\alpha}$	$p'^{1/\alpha} \left(\frac{T}{T_{ce}}\right)^\beta$	$p'^{1/\alpha} \left(\frac{T_{ve}}{T}\right)^\beta$	$\frac{r'}{0.35}$	$\left(\frac{T_{vle}}{T}\right)^\beta$
RDU	$w^+(p)$	$w^+(p) \left(\frac{T}{T_{ce}}\right)^\beta$	$w^+(p) \left(\frac{T_{ve}}{T}\right)^\beta$	$\frac{w^+(r')}{w^+(0.35)}$	$\left(\frac{T_{vle}}{T}\right)^\beta$
DA	$\frac{p}{1+(1-p)\delta}$	$\frac{p}{1+(1-p)\delta} \left(\frac{T}{T_{ce}}\right)^\beta$	$\frac{p}{1+(1-p)\delta} \left(\frac{T_{ve}}{T}\right)^\beta$	$\frac{r(1+0.65\delta)}{0.35(1+(1-r)\delta)}$	$\left(\frac{T_{vle}}{T}\right)^\beta$
PT	A	<i>RP Death:</i> $w^+(p') \left(\frac{T}{T_{ce}}\right)^\beta$	$A \left(\frac{T_{ve}}{T}\right)^\beta$	<i>RP Death:</i> $\frac{w^+(r')}{w^+(0.35)}$	<i>RP Death:</i> $\left(\frac{T_{vle}}{T}\right)^\beta$
		<i>RP (FH, T):</i> $(1-w^-(1-p')) \left(\frac{T}{T_{ce}}\right)^\beta$		<i>RP (Q, T)</i> B	<i>RP (Q, T)</i> $\left(\frac{T_{vle}}{T}\right)^\beta$
				<i>RP (FH, T)</i> $\frac{1-w^-(1-r)}{1-w^-(0.65)}$	

Note: RP stands for reference point. $A = \frac{w^+(p)}{w^+(p)+\lambda w^-(1-p)}$. $B = \frac{w^+(r)}{w^+(r)+\lambda(w^-(1-r)-w^-(0.65))}$

We then dropped the assumption of expected utility and examined the impact of evaluating the five methods under prospect theory, rank-dependent utility, disappointment aversion, and the two gambling effect models assuming power utility for life duration for all

models. To operationalize Diecidue et al.'s (2004) model, we assumed that $V(Q,T) = \alpha U(Q,T)$. We also tried several other specifications, but this specification yielded the best fit among those that ensure that $V(\text{death}) = 0$. The benefits of certainty are positive when $\alpha > 1$. This implies that utilities by the riskless-risk methods exceed those by the risk-risk methods. To operationalize the model of Bleichrodt and Schmidt (2002), we assumed that $V(Q,T) = (U(Q,T))^\alpha$. We normalized U and V on $[0,1]$ which implies that V is more concave than U when $\alpha < 1$. This implies, in turn, that health state utilities by the riskless-risk methods exceed those by the risk-risk methods. The third row of Table 2 shows $H(Q)$ under Diecidue et al.'s (2004) (DSW) model; the fourth row under Bleichrodt and Schmidt's (2002) (BS) model.

To be able to evaluate the methods under prospect theory we had to specify for each method the location of the reference point. Here we followed the suggestions of Hershey and Schoemaker (1985) and Bleichrodt et al (2001) and took (Q,T) as the reference point in the PE and in the VE and either (FH,T) or Death in the CE. In the PLE we took either (FH,T) , or (Q,T) or Death as the reference point. In the VLE, (FH,T_{vle}) is implausible as a reference point and we only analyzed the data for reference points (Q,T) or Death. The final row of Table 2 shows the utilities under prospect theory with the different reference points. To compute the probability weights, we assumed Eq. (6).

5. Experiment

Background

Subjects were sixty-five economics students (aged between 22 and 29) from the University of Murcia. They were paid €36 to participate in five experimental sessions, each lasting approximately one hour. In each experimental session a different method was elicited. The experiment was carried out in small group sessions with at most six subjects per group. The sessions were separated by at least one week. Prior to the actual experiment, the questionnaire

was tested in several pilot sessions using university staff as subjects.

We elicited the utility of two health states. The health states were described through the EuroQol system, a widely used instrument to describe health states in medical research. The description of the health states is given in Table 3. Throughout the experiment, the health states were anonymously labeled as A and B to avoid framing bias. Preferences were elicited through a sequence of choices. Empirical evidence suggests that choice-based procedures lead to more consistent data than matching procedures (Bostic et al., 1990).

Table 3: The description of health states A and B

Health state A	Health state B
<ul style="list-style-type: none"> • Some problems walking about • Some problems performing self-care activities (e.g. eating, washing, dressing) • No problems performing usual activities (e.g. work, study, family or leisure activities) • Moderate pain or discomfort • Moderately anxious or depressed 	<ul style="list-style-type: none"> • Some problems walking about • Some problems performing self-care activities (e.g. eating, washing, dressing) • Unable to perform usual activities (e.g. work, study, family or leisure activities) • Moderate pain or discomfort • Moderately anxious or depressed

Details

Recruitment of subjects took place one week before the actual experiment started. At the recruitment, subjects received information about the experiment and were asked to read the descriptions of the two health states. In addition, subjects were handed a practice question on the PE method. They were asked to answer this practice question at home. This procedure intended to familiarize subjects with the PE method. Prior to the start of the first experimental session, during which the PE method was administered, subjects were asked to explain their answer to the practice question. When we were not convinced that a subject understood the task, we explained

it again until we were convinced that he understood the task. The same procedure was repeated for each of the remaining experimental sessions. Subjects received a practice question to take home of the method that would be administered in the next session, and before the actual experiment started they had to explain their answer to the question.

The order in which the methods were administered was: first session PE, second CE, third PLE, fourth VE, and fifth VLE. The experiment was part of a larger experiment. The presence of the other experimental tasks and the delay of at least one week between the sessions made it unlikely that subjects would recall their previous answers and would note the relationship between the sessions.²

At the beginning of each experimental session, instructions were read aloud and an additional practice question was given. We asked six questions per method by combining each health state (A and B) with three values for the gauge duration T: 13, 24, and 38 years. We used life durations substantially lower than subjects' life-expectancy to avoid perception problems: subjects may find it hard to perceive living for very long durations which exceed their life-expectancy. To avoid order effects, we varied the order in which the different questions were asked. So as to minimize response errors, subjects had to confirm the elicited indifference value after each question.

We also determined, through a choice-based procedure, the life duration T' that made a subject indifferent between T years in health state A and T' years in full health, for T equal to 13, 24, and 38 years. The same question was asked for health state B. These questions were included to test the appropriateness of assuming power utility for life duration. Under power utility, we should find that the ratio $\frac{T'}{T}$ is constant. Also, in the presence of multiplicativity, the condition that the ratio $\frac{T'}{T}$ is constant implies that the utility for life duration must be a power function

(Doctor and Miyamoto 2003). These questions were asked in the second experimental session.

Analysis

As normality was rejected by the Kolmogorov-Smirnov-Lilliefors test, we tested significance of differences among the five methods by the nonparametric Friedman ranks sum test. When the hypothesis of equality was rejected by the Friedman test, we performed multiple pairwise comparisons by the Wilcoxon signed rank test to examine differences between each pair of methods. We used a significance level of 1% in the Wilcoxon tests to control for experimentwise Type I error, the phenomenon that when many tests on a given level are performed, some will be significant by coincidence (multiple significance testing). The Friedman test was also used to test for the appropriateness of assuming multiplicativity (p' should be the same in the three PE questions and r' should be the same in the three PLE questions) and the power function ($\frac{T'}{T}$ should be constant, see above).

A distribution-free algorithm was used to determine the optimal values of the relevant parameters in the different theories. We started by setting the relevant parameter(s) equal to one and then we searched for the values of the parameters that minimized the sum of squared differences between the elicited utilities.³ We varied β between 0.05 and 2, γ^+ and γ^- between 0.25 and 2, λ between 0.25 and 4, δ between -1 and 10, α in Diecidue et al's (2004) and in Bleichrodt and Schmidt's (2002) gambling effect model between 0.05 and 2. Using wider bounds caused the program in some cases to choose extreme and implausible values so that all utilities were close to zero. The range of parameters used includes the estimates from the existing empirical literature. The optimal parameters were determined with an accuracy of 0.01.

² An example of a PE question can be found on www.bmg.eur.nl/personal/bleichrodt/FormulationPE.pdf. The wording of the other questions was similar.

³ To examine the sensitivity of the results to outliers, we also determined the parameters that minimized the sum of absolute differences. There were only small differences between the two sets of estimates.

6. Results

Two subjects were excluded from the analyses of health state A and 19 from the analyses of health state B, because their choices implied that they did not always prefer more life-years to less. This left 63 and 46 subjects in the analyses of health states A and B, respectively. More subjects had to be excluded for health state B, because B is a worse health state than A. The worse a health state, the more likely there is a duration for which subjects do not prefer additional life-years. The subjects who were excluded were those with the lowest utilities. The fact that more subjects were excluded for health state B than for health state A will not bias our conclusions, because these are not based on comparisons between the utilities for health states A and B.

Table 4: Median responses. Interquartile ranges in square brackets.

Health state A						
T	PE	CE	VE	PLE	VLE	
13y.	0.58 (0.54-0.63)	17 (14-22)	17 (11-21)	0.16 (0.12-0.19)	6 (4-6)	
24y.	0.68 (0.65-0.73)	27 (23-35)	28 (22-36)	0.17 (0.12-0.19)	11 (9-13)	
38y.	0.72 (0.68-0.77)	42 (38-48)	46 (41-48)	0.17 (0.13-0.18)	18 (16-19)	
Health state B						
T	PE	CE	VE	PLE	VLE	
13y.	0.49 (0.45-0.55)	18 (13-25)	14 (10-18)	0.13 (0.11-0.16)	4 (3-6)	
24y.	0.57 (0.51-0.65)	31 (24-36)	21 (14-38)	0.13 (0.11-0.16)	8 (7-11)	
38y.	0.59 (0.50-0.68)	44 (39-48)	43 (32-51)	0.13 (0.11-0.14)	14 (12-17)	

Table 4 shows the median responses and in square brackets the interquartile range of the responses. The mean responses were similar to the medians. Because health state A is better than health state B, we should observe higher responses in the PE, VE, PLE, and VLE methods and

lower responses in the CE methods for health state A. All subjects satisfied this consistency requirement. Note that r' was much lower than 0.35, so probability weighting could affect the PLE questions.

Contrary to the predictions of expected utility, rank-dependent utility, disappointment aversion, and the two gambling effect models, we could reject equality of T , T_{ce} , and T_{ve} for both health states and for all three gauge durations ($P < 0.01$). The finding that T_{ce} significantly exceeded T ($P < 0.001$ in all cases) is consistent under prospect theory with more risk aversion for mixed prospects. Contrary to prospect theory, we also found that T_{ve} generally exceeded T with the exception of the second question for health state B. The difference was always significant for health state A ($P < 0.001$). For health state B the difference was marginally significant for the first and third question ($P = 0.035$ and $P = 0.025$) and not significant for the second question ($P = 0.650$).

The tests of multiplicativity yielded mixed results. For both health states we could reject the hypothesis that the probabilities in the three PE questions were equal ($P < 0.01$). However, we could not reject the hypothesis that the probabilities in the three PLE questions were equal ($P = 0.148$ for health state A and $P = 0.085$ for health state B). The tests of the appropriateness of using power utility for life duration yielded positive results: for both health states we could not reject the hypothesis that the ratio $\frac{T'}{T}$ was constant ($P = 0.565$ for health state A, $P = 0.085$ for health state B).

Figure 1 shows the median utilities under expected utility with linear utility for life duration. For both health states and for all gauge durations we found significant differences between the five methods ($P < 0.001$ in all cases). The typical pattern was $VE > PE > CE > PLE > VLE$. The differences between PLE and VLE were, however, not significant. With few exceptions, all other paired differences were significant.

The systematic differences between the methods could not be explained by the assumed linearity of the utility for life duration: Expected utility with power utility for life duration could not explain the results either. The first line of Table 5 shows the medians of the optimal power coefficients for each gauge duration and for each health state. Under expected utility with these optimal power coefficients, we could clearly reject the hypothesis that the five methods yield the same utilities ($P < 0.001$ for both health states and for all gauge durations). Contrary to expected utility with linear utility, all differences between PLE and VLE were significant.

Figure 1: Median utilities under expected utility with linear utility for life duration

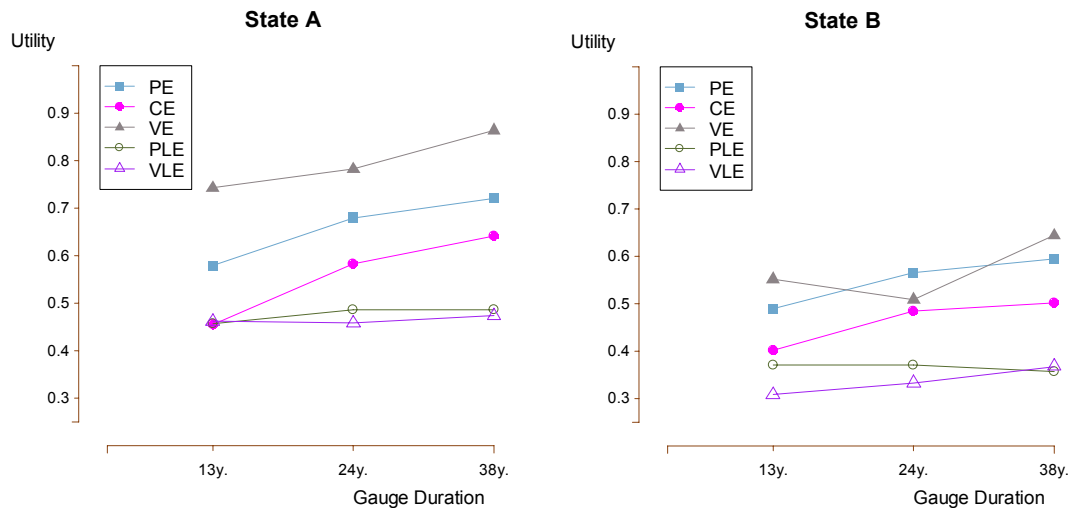


Table 5 also shows the optimal parameters for the other models we studied. The parameter estimate for α in Diecidue et al.'s (2004) gambling effect model exceeds one which indicates that the benefit of certainty was positive and that the function V exceeded the function U. The finding that $\alpha < 1$ in Bleichrodt and Schmidt's (2002) gambling effect model implies, as expected, that in their model the function V was more concave than the function U. The obtained estimates suggest that the difference in concavity was substantial. We found a similar degree of probability weighting under rank-dependent utility as other studies using health outcomes (Bleichrodt and Pinto 2000). The estimates for Gul's theory of disappointment aversion indicate

strong degrees of disappointment aversion. The final lines of Table 5 display the optimal parameter estimates for prospect theory where the reference point in the CE, PLE, and VLE methods is death. We also analyzed the data under the other possible reference points, but these did not lead to smaller differences between the methods and the convergence and fit at the individual level was far worse. Compared with other studies, we observed less probability weighting and comparable loss aversion. The findings on utility curvature varied with the theory examined: from strongly curved under expected utility and rank-dependent utility to moderately curved under prospect theory to slightly curved under the two gambling effect models to slightly convex (in four out of six cases) under disappointment aversion.

Table 5: Parameter estimates

Model	Duration	Health state A			Health state B		
		13	24	38	13	24	38
<i>EU-power model</i>							
β		0.46	0.45	0.48	0.60	0.53	0.60
<i>Gambling-effect models</i>							
Diecidue et al.							
α		1.26	1.56	1.56	1.53	1.55	1.66
β		0.60	0.80	0.93	0.89	0.80	0.93
Bleichrodt & Schmidt							
α		0.68	0.48	0.43	0.64	0.58	0.55
β		0.63	0.85	0.93	0.90	0.84	0.96
<i>Prospect Theory</i>							
β		0.86	0.73	0.65	0.78	0.65	0.65
λ		2.13	2.00	1.84	1.83	1.77	1.53
γ^+		0.93	0.79	0.73	0.95	0.84	0.77
γ^-		0.80	0.80	0.80	0.90	1.20	0.80
<i>Rank-dependent utility</i>							
γ		0.79	0.79	0.79	0.79	0.74	0.73
β		0.49	0.49	0.49	0.58	0.53	0.60
<i>Disappointment av.</i>							
δ		1.20	3.10	3.80	2.40	2.30	2.90
β		0.84	1.03	1.15	1.10	0.91	1.14

We also examined the data under prospect theory with the probability weighting and loss

aversion parameters obtained by Tversky and Kahneman (1992): $\gamma^+ = 0.61$, $\gamma^- = 0.69$, and $\lambda = 2.25$. We analyzed this case because Bleichrodt et al. (2001) were able to remove all discrepancies between PE and CE using Tversky and Kahneman's values.

None of the nonexpected utility theories could fully explain the data. The differences between the five methods were significant for all theories ($P < 0.01$) except for health state B under prospect theory with the optimal parameters where the differences between the methods were only marginally significant ($P = 0.040$, $P = 0.036$, and $P = 0.034$ for gauge durations 13 years, 24 years, and 38 years, respectively).

Table 6: Number of Significant Pairwise Differences Between Methods

Model	Duration	Health state A			Health state B		
		13	24	38	13	24	38
EU-linear		7	7	9	6	7	8
EU-power model		9	8	9	6	6	7
Diecidue et al.		7	7	6	5	5	4
Bleichrodt & Schmidt		8	6	5	4	2	2
Rank-Dep. Utility		10	8	8	8	3	3
Disappointment Av.		6	7	7	6	5	4
Prospect Theory TK		6	7	6	4	4	6
Prospect Theory Opt.		3	4	4	1	2	1

Note: Prospect Theory TK stands for Prospect Theory with the probability weighting and loss aversion parameters obtained by Tversky and Kahneman (1992)

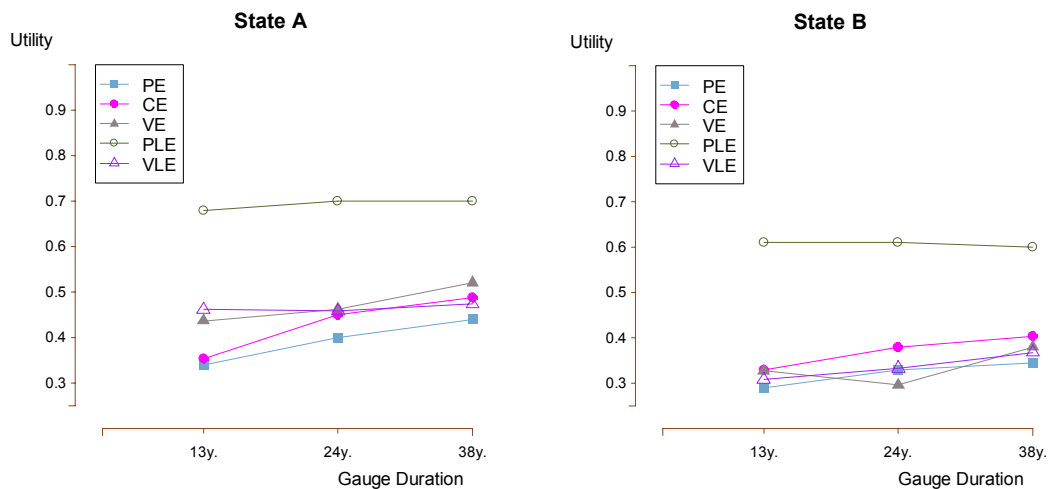
Table 6 shows the number of significant pairwise differences between the five elicitation methods. For each gauge duration, there were 10 comparisons between methods. The table shows that using power utility instead of linear utility did not improve the performance of expected utility, that many differences remained significant under prospect theory with the probability weighting and loss aversion parameters obtained by Tversky and Kahneman (1992)⁴, and that the number of significant differences was lowest under prospect theory with the optimal parameters.

⁴ Data in the Table are under linear utility for life duration. Similar data obtained when we allowed for curved utility for life duration.

The number of significant differences was also relatively low under Bleichrodt and Schmidt's gambling effect model.

Figure 2⁵ shows the results under prospect theory with Tversky and Kahneman's (1992) values and linear utility for life duration. The figure shows that the main problem is that PLE was too high. As can be seen from Table 2, this problem cannot be solved by allowing for utility curvature because neither the PE nor the PLE are affected by the assumed form of the utility for life duration and, hence, their discrepancy will remain after correction for utility curvature. The difference between the PLE and the other methods can under prospect theory only be explained by a difference in the degree of probability weighting in our study as compared with Tversky and Kahneman (1992). In particular, given that the response r' in the PLE was generally between 0.10 and 0.20, our data suggest less overweighting of probabilities in that range than suggested by Tversky and Kahneman (1992) assuming that there is less probability weighting around 0.35.

Figure 2: Median utilities under prospect theory with Tversky and Kahneman's values and linear utility for life duration

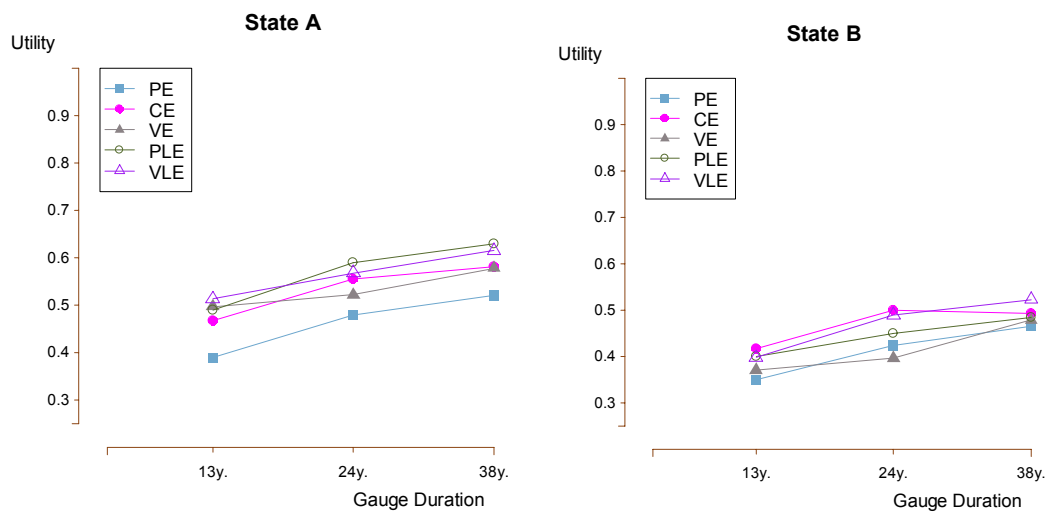


⁵ Figures for the other cases can be found on www.bmg.eur.nl/personal/bleichrodt/ExtraFigures.pdf.

Figure 2 also shows that prospect theory could rather well explain the differences between the other four methods, except that for health state A the PE was too low. In these methods, the probabilities involved generally exceeded 0.35, suggesting comparable probability weighting in the range $[0.35,1]$ as found by Tversky and Kahneman (1992). The finding that CE generally exceeds PE is in contrast with Bleichrodt et al. (2001). The difference between PE and CE was significant for health state A and gauge durations 24 years and 38 years and for health state B and gauge duration 38 years.

Finally, Figure 3 shows the median utilities under prospect theory with the optimal parameters. In general the utilities were close; the exception is that for health state A the PE was too low. For health state B there seem to be no systematic differences between the five methods.

Figure 3: Median utilities under prospect theory with optimal parameters



7. Discussion

Our findings confirm that methods that according to expected utility are equivalent, produce systematically different results. The data suggest that evaluating riskless-risk methods by expected utility leads to utilities that are too high. Of the alternatives for expected utility that we

examined, no single model could explain the data entirely. The findings that under expected utility there was no significant difference between risk-risk methods (when the utility for duration is close to linear) and that the utilities obtained by riskless-risk methods were higher than those obtained by risk-risk methods agree with the predictions of the gambling effect models. The gambling effect models cannot, however, explain the systematic discrepancies between the riskless-risk methods. These discrepancies can be resolved by prospect theory, the theory that was most consistent with the data. The finding that the two risk-risk methods yielded comparable results under expected utility is, however, harder to reconcile with prospect theory. As we argued before, one explanation for this finding may be that there was little overweighting of small probabilities in our study. Anyway, our data seem to add to the evidence that violations of expected utility primarily occur when one of the prospects under evaluation is riskless.

That said, it may be too optimistic to assume that one single model could explain all data. After all, the number of deviations from expected utility that we considered was limited and other biases will most likely have affected people's responses. In particular, we ignored the impact of scale compatibility, which predicts that the attention people devote to an attribute is higher as this attribute is more compatible with the response scale used (Tversky et al. 1988, Delqu   1993). Scale compatibility could explain for instance why T_{ve} , the response to the VE question, generally exceeds T , a finding that none of the theories considered could explain. The reason we did not include scale compatibility was that its impact is ambiguous. For instance, in the PE method, where the probability p was determined that made a subject indifferent between the risky prospect ($p:(FH,T); \text{Death}$) and the riskless prospect (Q,T), the response scale is probability and scale compatibility predicts that consequently people will focus on probability when evaluating prospects. However, people could either focus on the probability p of the good outcome full health or on the probability $1-p$ of the bad outcome Death and the direction of the bias due to scale compatibility depends on which probability people focus on.

As noted before, we focused on health for reasons of practical relevance. A drawback of using health outcomes is that we had to assume multiplicativity to be able to compute utilities. We observed mixed evidence on multiplicativity. It should be noted that even when multiplicativity is violated we can still compare the PE utilities with the PLE utilities. In that case we compare the utilities $U(Q,T)$ under PE and PLE rather than the utilities $H(Q)$. Our conclusions were not affected when we only compared the PE with the PLE. Thus the assumption of multiplicativity does not seem to be critical in the findings of our study.

To operationalize prospect theory, we had to make assumptions about the location of the reference point. Empirical evidence seems to support our assumptions about the location of the reference point in the riskless-risk methods. No evidence exists about the location of the reference point in the risk-risk questions. Here we extended the arguments of Hershey and Schoemaker (1985) and Bleichrodt et al. (2001). Possibly, however, the formation of a reference point is more complicated when both prospects are risky. For example, people may take a risky prospect as their reference point rather than a single outcome. It might be that in the PLE method, where the probability r was determined that made a subject indifferent between the risky prospects ($r: (FH,T); \text{Death}$) and ($p: (Q,T); \text{Death}$)), the reference point for the best outcome of the prospects was (Q,T) , but for the worst outcomes it was Death. It is not clear how to cover such a situation. In particular, it is not clear how to model probability weighting when the reference point is a risky prospect. Sugden (2003) presented an extension when people do not weight probabilities.

The use of students as subjects may limit the generalizability of our findings. Empirical evidence on health utility measurement has suggested, however, that there are 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).

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 tasks that we asked our subjects to perform, there appears to be no systematic difference in the general pattern of responses, although real incentives tend to reduce data variability.

The order in which the five methods were administered was fixed. We found no evidence that this affected the results. For example, the standard deviations of individual responses in the VE were comparable to those in the CE even though VE was administered in a later session. This finding contradicts the hypothesis that people became more aware of their true preferences over the sessions leading to lower response error. Also, we found no evidence that people behaved more in agreement with expected utility in later sessions, contradicting the hypothesis that more experience with choosing between prospects may lead to fewer violations of expected utility.

Let us finally discuss the implications of our research for practical decision analysis. Our findings suggest that the common practice in medical decision analysis to measure utilities under risk by probability equivalence or certainty equivalence methods and to evaluate the responses by expected utility will lead to biased utilities; our data suggest that this bias is upwards. Instead, riskless-risk methods, like the probability equivalence and certainty equivalence methods, should be evaluated through prospect theory. For riskless-risk methods, the parameters found by Tversky and Kahneman (1992) performed well. A possible alternative for using riskless-risk methods and prospect theory may be to use risk-risk methods and expected utility. We found no inconsistencies when evaluating risk-risk methods by expected utility. As we argued, this can be explained by the virtual absence of overweighting of probabilities in the interval $[0.10, 0.20]$. To what extent our findings can be generalized to other stimuli and outcome domains remains an open question. The preliminary findings in McCord and deNeufville (1986) for money agree with our findings, which gives some confidence that comparable findings may be obtained for other stimuli and outcome domains.

Appendix: Derivation of the formulas in Table 2.

Throughout we assume multiplicativity and power utility for life duration $U(Q,T) = H(Q)T^\beta$ and we use the scaling $H(FH) = 1$ and $U(\text{Death}) = 0$. In Bleichrodt and Schmidt's (2002) gambling effect model, we need to normalize utilities on $[0,1]$. Hence, we set $L(38) = 1$ in that model. Expected utility with linear utility is the special case of expected utility with power utility where $\beta = 1$. Disappointment aversion is the special case of rank-dependent utility where $w^+(p) = \frac{p}{1+(1-p)\delta}$.

PE method

The indifference $(Q,T) \sim ((FH,T), p'; \text{Death})$ yields under expected utility $H(Q)T^\beta = p'T^\beta$ or $H(Q) = p'$. Under Diecidue et al.'s (2004) gambling effect model, we obtain $\alpha(H(Q)T^\beta) = p'T^\beta$. Rearranging gives $H(Q) = \frac{p'}{\alpha}$. Bleichrodt and Schmidt's (2002) gambling effect model gives $(H(Q)(T/38)^\beta)^\alpha = p'((T/38)^\beta)^\alpha$, or $H(Q) = (p')^{1/\alpha}$. Under rank-dependent utility we obtain $H(Q)T^\beta = w^+(p)T^\beta$ and, hence, $H(Q) = w^+(p)$. Under prospect theory, (3) yields $H(Q)T^\beta = H(Q)T^\beta + w^+(p')(T^\beta - H(Q)T^\beta) - \lambda w^-(1-p')H(Q)T^\beta$. Rearranging and deleting the common term T^β gives $H(Q) = \frac{w^+(p)}{w^+(p) + \lambda w^-(1-p)}$.

CE method

The indifference $(Q, T_{ce}) \sim ((FH, T), p'; \text{Death})$ yields under expected utility $H(Q)T_{ce}^\beta = p'T^\beta$ and thus $H(Q) = p'(\frac{T}{T_{ce}})^\beta$. Under Diecidue et al.'s (2004) gambling effect model, we obtain $\alpha(H(Q)T_{ce}^\beta) = p'T^\beta$. Rearranging gives $H(Q) = \frac{p'}{\alpha}(\frac{T}{T_{ce}})^\beta$. Bleichrodt and Schmidt's (2002)

gambling effect model gives $(H(Q)(T_{ce}/38)^\beta)^\alpha = p'((T/38)^\beta)^\alpha$, or $H(Q) = (p')^{1/\alpha} (\frac{T}{T_{ce}})^\beta$. Under rank-dependent utility, $H(Q)T_{ce}^\beta = w^+(p')T^\beta$ and thus $H(Q) = w^+(p')(\frac{T}{T_{ce}})^\beta$. The same expression holds for prospect theory with reference point death. If the reference point is (FH,T), (5) gives $T^\beta - \lambda(H(Q)T_{ce}^\beta - T^\beta) = T^\beta - \lambda w^-(1-p)(-T^\beta)$. Deleting common terms and rearranging gives $H(Q) = (1-w^-(1-p'))(\frac{T}{T_{ce}})^\beta$.

VE method

The indifference $(Q,T) \sim ((FH,T_{ve}), p'; \text{Death})$ yields under expected utility $H(Q)T^\beta = p' T_{ve}^\beta$ and thus $H(Q) = p'(\frac{T_{ve}}{T})^\beta$. Under Diecidue et al.'s (2004) gambling effect model, we obtain $\alpha(H(Q)T^\beta) = p' T_{ve}^\beta$. Rearranging gives $H(Q) = \frac{p'}{\alpha} (\frac{T_{ve}}{T})^\beta$. Bleichrodt and Schmidt's (2002) gambling effect model gives $(H(Q)(T/38)^\beta)^\alpha = p'((T_{ve}/38)^\beta)^\alpha$, or $H(Q) = (p')^{1/\alpha} (\frac{T_{ve}}{T})^\beta$. Under rank-dependent utility, $H(Q)T^\beta = w^+(p)T_{ve}^\beta$, or $H(Q) = p'(\frac{T_{ve}}{T})^\beta$. Prospect theory with reference point (Q,T) gives by (3), $H(Q)T^\beta = H(Q) T^\beta + w^+(p')(\frac{T_{ve}}{T})^\beta - H(Q)T^\beta - \lambda w^-(1-p')$
 $H(Q)T^\beta$. Rearranging gives $H(Q) = \frac{w^+(p)}{w^+(p)+\lambda w^-(1-p)} (\frac{T_{ve}}{T})^\beta$.

PLE method

The indifference $((Q,T), 0.35; \text{Death}) \sim ((FH,T), r'; \text{Death})$ yields under expected utility $0.35H(Q)T^\beta = r'T^\beta$, or $H(Q) = \frac{r'}{0.35}$. The two gambling effect models agree with expected utility when both prospects are risky. Under rank-dependent utility, $w^+(0.35)H(Q)T^\beta = w^+(r')T^\beta$, or $H(Q) = \frac{w^+(r)}{w^+(0.35)}$. Prospect theory with reference point Death is identical to rank-dependent utility. If the reference point is (Q,T), (3) and (5) give $H(Q)T^\beta - \lambda w^-(0.65) H(Q)T^\beta = H(Q)T^\beta + w^+(r')(T$

$\beta - H(Q)T^\beta - \lambda w^-(1-r) H(Q)T^\beta$. Rearranging gives $H(Q) = \frac{w^+(r)}{w^+(r) + \lambda(w^-(1-r) - w^-(0.65))}$. If the

reference point is (FH,T), (5) gives $T^\beta - \lambda w^-(1-r)T^\beta = T^\beta -$

$\lambda w^-(0.65)T^\beta - \lambda(1-w^-(0.65))(T^\beta - H(Q)T^\beta)$. Rearranging gives $H(Q) = \frac{1-w^-(1-r)}{1-w^-(0.65)}$.

VLE method

The indifference $((Q,T), 0.35; \text{Death}) \sim ((FH, T_{vle}), 0.35; \text{Death})$ yields under expected utility $0.35H(Q)T^\beta = 0.35 T_{vle}^\beta$, or $H(Q) = (\frac{T_{vle}}{T})^\beta$. The two gambling effect models agree with expected utility when both prospects are risky. Rank-dependent utility gives $w^+(0.35)H(Q)T^\beta = w^+(0.35)T_{vle}^\beta$ or $H(Q) = (\frac{T_{vle}}{T})^\beta$. The same expression holds for prospect theory with reference point Death. If the reference point is (Q,T), (3) and (5) give $H(Q)T^\beta - \lambda w^-(0.65) H(Q)T^\beta = H(Q)T^\beta + w^+(0.35)(T_{vle}^\beta - H(Q)T^\beta) - \lambda w^-(0.65) H(Q)T^\beta$. Deleting common terms and rearranging gives $H(Q) = (\frac{T_{vle}}{T})^\beta$. If the reference point is (FH,T) then (5) gives $T_{vle}^\beta - \lambda w^-(0.65)T_{vle}^\beta = T_{vle}^\beta - \lambda w^-(0.65)T_{vle}^\beta - \lambda(1-w^-(0.65))(T_{vle}^\beta - H(Q)T^\beta)$. Deleting common terms and rearranging gives $H(Q) = (\frac{T_{vle}}{T})^\beta$.

References

- Abdellaoui, M. 2000. Parameter-free elicitation of utilities and probability weighting functions. *Management Science* **46** 1497-1512.
- Bleichrodt, H. 2001. Probability weighting in choice under risk: An empirical test. *Journal of Risk and Uncertainty* **23** 185-198.
- Bleichrodt, H., J. L. Pinto. 2000. A parameter-free elicitation of the probability weighting function in medical decision analysis. *Management Science* **46** 1485-1496.

- Bleichrodt, H., J. L. Pinto. 2004. The validity of QALYs under non-expected utility. *The Economic Journal* (in press).
- Bleichrodt, H., J. L. Pinto, P. P. Wakker. 2001. Using descriptive findings of prospect theory to improve the prescriptive use of expected utility. *Management Science* **47** 1498-1514.
- Bleichrodt, H., U. Schmidt. 2002. A context-dependent model of the gambling effect. *Management Science* **48** 802-812.
- Camerer, C. 1992. Recent tests of generalizations of expected utility theory. W. Edwards, eds. *Utility : Theories, measurement and applications*. Kluwer Academic Publishers, Boston, M.A., 207-251.
- Cher, D. J., J. Miyamoto, L. A. Lenert. 1997. Incorporating risk attitude into Markov-process decision models. *Medical Decision Making* **17** 340-350.
- Conlisk, J. 1989. Three variants on the Allais paradox. *American Economic Review* **79** 392-407.
- de Wit, G. A., J. J. van Busschbach, F. T. de Charro. 2000. Sensitivity and perspective in the valuation of health status. *Health Economics* **9** 109-126.
- Delquié, P. 1993. Inconsistent trade-offs between attributes: New evidence in preference assessment biases. *Management Science* **39** 1382-1395.
- Diecidue, E., U. Schmidt, P. P. Wakker. 2004. The utility of gambling reconsidered. *Journal of Risk and Uncertainty* **29** 241-259.
- Doctor, J. N., H. Bleichrodt, J. Miyamoto, N. R. Temkin, S. Dikmen. 2004. A new and more robust test of QALYs. *Journal of Health Economics* **23** 353-367.
- Doctor, J. N., J. Miyamoto. 2003. Deriving quality-adjusted life-years (QALYs) from constant proportional time tradeoff and risk posture conditions. *Journal of Mathematical Psychology* **47** 557-567.
- Gonzalez, R., G. Wu. 1999. On the form of the probability weighting function. *Cognitive Psychology* **38** 129-166.
- Gul, F. 1991. A theory of disappointment aversion. *Econometrica* **59** 667-686.
- Harless, D., C. F. Camerer. 1994. The predictive utility of generalized expected utility theories. *Econometrica* **62** 1251-1289.

- Hershey, J. C., P. J. H. Schoemaker. 1985. Probability versus certainty equivalence methods in utility measurement: Are they equivalent? *Management Science* **31** 1213-1231.
- Hertwig, R., A. Ortmann. 2001. Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences* **24** 383-451.
- Kahneman, D., A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* **47** 263-291.
- Keller, R. L., D. N. Kleinmuntz. 1998. Is this the right time for a new decision analysis journal? *Decision Analysis Society Newsletter* **17**
- Llewellyn-Thomas, H., H. J. Sutherland, R. Tibshirani, A. Ciampi, J. E. Till, N. F. Boyd. 1982. The measurement of patients' values in medicine. *Medical Decision Making* **2** 449-462.
- McCord, M., R. de Neufville. 1986. Lottery equivalents: Reduction of the certainty effect problem in utility assessment. *Management Science* **32** 56-60.
- Miyamoto, J. M., S. A. Eraker. 1985. Parameter estimates for a QALY utility model. *Medical Decision Making* **5** 191-213.
- Miyamoto, J. M., S. A. Eraker. 1988. A multiplicative model of the utility of survival duration and health quality. *Journal of Experimental Psychology: General* **117** 3-20.
- Morrison, G. C. 2000. The endowment effect and expected utility. *Scottish Journal of Political Economy* **47** 183-197.
- Oliver, A. 2003. The internal consistency of the standard gamble: Tests after adjusting for prospect theory. *Journal of Health Economics* **22** 659-674.
- Pliskin, J. S., D. S. Shepard, M. C. Weinstein. 1980. Utility functions for life years and health status. *Operations Research* **28** 206-223.
- Quiggin, J. 1981. Risk perception and risk aversion among Australian farmers. *Australian Journal of Agricultural Economics* **25** 160-169.
- Robinson, A., G. Loomes, M. Jones-Lee. 2001. Visual analog scales, standard gambles, and relative risk aversion. *Medical Decision Making* **21** 17-27.

- Rutten-van Mólken, M. P., C. H. Bakker, E. K. A. van Doorslaer, S. van der Linden. 1995. Methodological issues of patient utility measurement. Experience from two clinical trials. *Medical Care* **33** 922-937.
- Shalev, J. 2000. Loss aversion equilibrium. *International Journal of Game Theory* **29** 269-287.
- Smith, J. E., D. von Winterfeldt. 2004. Decision analysis in *Management Science*. *Management Science* **50** 561-574.
- Stalmeier, P. F. M., T. G. G. Bezembinder. 1999. The discrepancy between risky and riskless utilities: A matter of framing? *Medical Decision Making* **19** 435-447.
- Starmer, C. 2000. Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature* **28** 332-382.
- Stiggelbout, A. M., G. M. Kiebert, J. Kievit, J. W. H. Leer, G. Stoter, J. C. J. M. de Haes. 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* **14** 82-90.
- Sugden, R. 2003. Reference-dependent subjective expected utility. *Journal of Economic Theory* **111** 172-191.
- Tversky, A., C. Fox. 1995. Weighing risk and uncertainty. *Psychological Review* **102** 269-283.
- Tversky, A., D. Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* **5** 297-323.
- Tversky, A., S. Sattath, P. Slovic. 1988. Contingent weighting in judgment and choice. *Psychological Review* **95** 371-384.
- Wu, G., R. Gonzalez. 1996. Curvature of the probability weighting function. *Management Science* **42** 1676-1690.