An elicitation of utility for quality of life under prospect theory

Attema, Arthur E.; Brouwer, Werner B.F.; I'Haridon, Olivier; Pinto Prades, Jose Luis

Published in:
Journal of Health Economics

DOI:
10.1016/j.jhealeco.2016.04.002

Publication date:
2016

Document Version
Author accepted manuscript

An elicitation of utility for quality of life under prospect theory[1]

Arthur E. Attema[a], Werner B.F. Brouwer[b], Olivier l'Haridon[c], and Jose Luis Pinto[d,e]

[a] (Corresponding author) iBMG, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands. E-mail: attema@bmg.eur.nl, --31-10.408.91.29 (O); --31-10.408.90.81 (F)
[b] iBMG, Erasmus University, Rotterdam, The Netherlands
[c] CREM, University of Rennes I, Rennes, France
[d] Department of Economics, University of Navarra, Pamplona, Spain
[e] Yunus Center for Social Business and Health, Glasgow Caledonian University, Pamplona, Spain

April, 2016

ABSTRACT. This paper performs several tests of decision analysis applied to the health domain. First, we conduct a test of the normative expected utility theory. Second, we investigate the possibility to elicit the more general prospect theory. We observe risk aversion for gains and losses and violations of expected utility. These results imply that mechanisms governing decisions in the health domain are similar to those in the monetary domain. However, we also report one important deviation: utility is universally concave for the health outcomes used in this study, in contrast to the commonly found S-shaped utility for monetary outcomes, with concave utility for gains and convex utility for losses.

Key Words: certainty equivalences, loss aversion, prospect theory, QALYs
JEL classification: I10

[1] This research was made possible through a grant from The Netherlands Organization for Health Research and Development (ZonMW), project number 152002041. We thank Han Bleichrodt and Peter P. Wakker for suggestions.

1. Introduction

Economic evaluations in health care are often based on cost-utility analyses, with health gains expressed in terms of Quality-Adjusted Life-Years (QALYs, Pliskin et al., 1980). Applications of the QALY model often assume expected utility (EU) theory to hold. However, it is by now well-known that EU has limited descriptive validity in many domains, including in health (Bleichrodt et al., 2007; Llewellyn-Thomas et al., 1982; Treadwell and Lenert, 1999).

Prospect Theory (PT, Tversky and Kahneman, 1992) has developed as an important alternative to EU, with more descriptive validity. Descriptively, we have to reckon with violations of EU and, hence, in order to measure utility, we should resort to descriptive models such as PT, even if we consider EU to be normative. PT's main deviations from EU are its reliance on a reference point, with people valuing deviations from this reference point rather than absolute outcomes, a steeper slope of utility for losses as seen from the reference point than for gains (*loss aversion*), as well as a nonlinear transformation of probabilities into decision weights (*probability weighting*).

Several studies have measured utility and/or probability weighting under PT, both for monetary outcomes (Abdellaoui, 2000; Bruhin et al., 2010) and for health outcomes (Bleichrodt and Pinto, 2000; Miyamoto and Eraker, 1989; Riddel and Kolstoe, 2013). Quantifications of loss aversion are increasingly performed as well. Abdellaoui et al. (2007; 2008; 2011) and Booij and van de Kuilen (2009) measured loss aversion in the monetary domain, whilst Attema et al. (2013), Bleichrodt and Pinto (2002) and Bleichrodt et al. (2007) measured loss aversion in the health domain. However, the latter studies also indicated that for health, the location of the reference point is not clear.

When combining PT with the QALY model, the latter is more complicated than many other decision models because it involves two attributes: longevity and quality of life (QoL) (Bleichrodt et al., 2009). An additional complication with this model is that QoL is a non-numerical attribute. Consequently, if we are to derive a full parameterization of the QALY model, we need to elicit values over both life duration and QoL. Although the utility for life duration has recently been elicited according to PT (Attema et al., 2013), no such attempt has yet been made for QoL. Abellán-Perpinán et al. (2009) applied PT to QoL estimations, but they did not measure PT's parameters.

This paper studies the influence of reference points in health systematically and proposes a complete elicitation of prospect theory for quality of life outcomes. Bleichrodt and Pinto (2002) and Bleichrodt et al. (2003) reported reference-dependency to be highly influential in the health domain, but left a quantification of the amount of loss aversion and a complete quantification of prospect theory for future research. In this paper, we generated reference points by explicit framing and investigated how their manipulation affected preferences towards quality of life.

The methodological objective of this research is to show how utility for health can be estimated using more realistic assumptions about the way in which subjects respond to elicitation

questions. Traditionally, it has been assumed that subjects respond to standard gamble questions using EU and the QALY model (Furlong et al., 2001). Since there is evidence that the traditional EU model is not descriptively valid, it is important to estimate QALYs using theories based on more realistic assumptions, such as PT. By applying a comprehensive method, we are able to measure utility for gains and losses, simultaneously with a loss aversion index and probability weighting. First, we show how utility would be estimated under the traditional assumption of EU. We proceed to show that these assumptions are incompatible with our evidence. More specifically, we conduct a direct test of traditional EU by investigating whether risk aversion is influenced by positioning of the reference point. Third, we show how to estimate utility using PT.

Interestingly, we find more risk aversion when the same outcomes are framed as losses than when they are framed as gains. This is in sharp contrast to findings of studies using monetary outcomes, where risk seeking behavior is commonly found in the loss domain (Tversky and Kahneman, 1992). Hence, our findings point to an important difference between health and money. The parameterization of PT reflects this difference by concave utility for losses and close to linear utility for gains. The evidence for probability weighting is mixed, with a slight tendency to pessimism for gains. We also observe loss aversion, to a similar extent as commonly found for monetary outcomes. In summary, traditional EU is found to be too constrained to explain behavior in health, suggesting that alternative theories accounting for sign-dependence are worthwhile.

In the next section we introduce notation and describe our measurement method. Section 3 presents details of the two experiments we performed, Section 4 summarizes the results, Section 5 summarizes and discusses the main findings, and Section 6 concludes.

2. Method

2.1. Decision models

The decision maker faces uncertainty about the health he will have in the future and chooses between binary prospects ($x_p y$) that give outcome x with probability p and outcome y with probability 1-p, with outcomes being health profiles. Our experiment only uses binary prospects and, hence, we do not consider prospects with more than two outcomes here. A decision maker's preference relation $\succcurlyeq$ is assumed to be a *weak order*, i.e., it is *transitive* and *complete*. The relation $\succ$ denotes the asymmetric part of $\succcurlyeq$, and ~ denotes indifference.

To represent the decision maker's preferences, we assume the generalized QALY model (Miyamoto and Eraker, 1988). This model evaluates health profiles x=(H,T) by the function $U(x)=V(H)\cdot W(T)$, with U(x) the total utility of a period T in the health state H, being a product of

V(H), the value of H, and W(T), the discounted weight of duration T. This paper focuses on ceteris paribus comparisons of health profiles with identical duration. Comparison of chronic health profiles hence reduces to comparison of QALY weights through V(H).

The generalized QALY model is compatible with several models of decision under uncertainty. If two health profiles, x and y, are possible, the traditional EU model evaluates preferences over these profiles by:

$$\mathrm{EU}(p, x; y) = p \cdot U(x) + (1 - p) \cdot U(y). \quad (1)$$

Because of the limited descriptive validity of EU, more general specifications that incorporate PT and rank-dependent utility have been developed in the health domain (Bleichrodt and Quiggin, 1997; Bleichrodt and Miyamoto, 2003). In PT, outcomes are defined as changes with respect to a reference outcome r rather than final health positions. An outcome x is a gain if $x \succcurlyeq r$ and a loss if $r \succ x$. A gain prospect contains only gains, a loss prospect contains only losses, and a mixed prospect contains a gain and a loss. According to PT, the evaluation of prospects becomes (Wakker, 2010)[2]:

$$\mathrm{PT}(p, x; y) = w^{+}(p) \cdot \big(U(x) - U(y)\big) + U(y) \quad (2a)$$

for gain prospects;

$$\mathrm{PT}(p, x; y) = w^{-}(p) \cdot \big(U(x) - U(y)\big) + U(y) \quad (2b)$$

for loss prospects, and;

$$\mathrm{PT}(p, x; y) = w^{+}(p) \cdot \big(U(x) - U(r)\big) + w^{-}(1 - p) \cdot \big(U(y) - U(r)\big) + U(r) \quad (2c)$$

for mixed prospects. In 2a-2c, $w^{+}(p)$ is a probability weighting function for gains and $w^{-}(p)$ is a probability weighting function for losses. Probability weighting functions are strictly increasing functions from [0,1] to [0,1] with $w^{i}(0)=0$ and $w^{i}(1)=1$, i=+/-. If $w^{+}(p)=p$ and $w^{-}(p)=p$, 2a-2c define a reference-dependent version of expected utility. Wakker (2010, Observation 7.11.1, p.231) shows that when there are only two outcomes, as in this study, PT for gains agrees with rank-dependent utility as well as many other theories. Thus, our utility measurement is even more general than PT for two outcomes.

In this paper, we do not consider qualitative health scales, but instead study quantitative health scores summarized as a percentage q of full health FH. The advantage of this representation is that we can measure it on a continuous scale. Other researches have used this approach and they have not

[2] Note that PT (Tversky and Kahneman, 1992) and original PT (Kahneman and Tversky, 1979) coincide for binary prospects, which we study in this paper.

reported important problems in the use of this scale (Baker et al., 2010; Dolan and Tsuchiya, 2009; Pinto-Prades et al., 2014).

Attema et al. (2013) applied the semi-parametric method proposed by Abdellaoui et al. (2008) to estimate EU's and PT's parameters in the context of life years. We use this method but apply it to QoL. Reference-dependent preferences can be more important in QoL because reference points are more subjective than in life expectancy. That is, in life expectancy the reference point could be the "expected life expectancy" at that age, but in QoL it is not that clear where the reference point is located. In addition to measuring the parameters, we also repeat the measurement using different reference points and compare the resulting parameter estimates, whereas Attema et al. (2013) considered only one reference point.

2.2. Elicitation of utility and decision weights

For each prospect j in the gain [loss] domain, we elicit a value $\bar{q}^j$ such that the respondent is indifferent between gaining [losing] $\bar{q}^j$ percentage points of QoL for certain and the prospect that provides a higher gain [loss] q_a^j with probability 0.5 and a lower gain [loss] q_b^j with probability 0.5. For any prospect, QoL deviations from full health are transitional, lasting for 1 year. Value $\bar{q}^j$ corresponds to the certainty equivalent (CE). In the gain domain, we have $q_a^j > q_b^j$, in the loss domain, $q_a^j < q_b^j$. Assuming the respondent's preferences can be represented by PT and a power utility function $V^+(q^j \cdot FH) = (q^j \cdot FH)^\alpha$, this indifference yields the following regression equation for gains[3]:

$$\bar{q}^j = \left(\omega^+ \cdot \left(\left(q_a^j\right)^\alpha - \left(q_b^j\right)^\alpha\right) + \left(q_b^j\right)^\alpha\right)^{1/\alpha} + e^j, \quad (3)$$

with $\omega^+ = w^+(0.5)$ and e^j an iid error term. This equation enables the simultaneous estimation of the utility parameter α and the weight ω^+ through nonlinear least squares. The procedure is similar in the loss domain with a utility $V^-(q^j \cdot FH) = -(-q^j \cdot FH)^\beta$ and $\omega^- = w^-(0.5)$, giving the following regression equation:

$$\bar{q}^j = -\left(\omega^- \cdot \left(\left(-q_a^j\right)^\beta - \left(-q_b^j\right)^\beta\right) + \left(-q_b^j\right)^\beta\right)^{1/\beta} + e^j. \quad (4)$$

[3] See Appendix A for a derivation of Eqs (3) and (4). The power utility used in the estimation was defined as $u(q)=(q+c)^\alpha - c^\alpha$ if $\alpha>0.01$, $u(q)=\log(q)$ if $\alpha<|0.01|$, $u(q)=-(q+c)^\alpha + c^\alpha$) if $\alpha<-0.01$, with the correction continuity parameter c set to 0.01. Wakker (2008, p.1336) has recommended the use of such correction continuity.

Under EU, for gains and losses indifferences yield Eq. (3) with the restriction $\omega^+ = p$.

2.3. Elicitation of loss aversion

In order to estimate a loss aversion index, we make use of the representation of Köbberling and Wakker (2005). The decision maker has a utility U which is defined by:

$$U(q^j) = \begin{cases} (q^j)^\alpha & \text{if } q^j \geq r \\ -\lambda \cdot (-q^j)^\beta & \text{if } q^j < r, \end{cases} \tag{5}$$

with loss aversion index λ and r the reference point. This model was also adopted by Tversky and Kahneman (1992).

The loss aversion index λ can be estimated by selecting an improvement [deterioration] of quality of life above [below] the reference point, and determining the loss [gain] below [above] the reference point for which the subject was indifferent between a treatment that with equal probability gives either a gain or a loss, and no treatment [staying at the reference point r].

Denoting q_a^j the gain, q_b^j the loss in the mixed prospect, and setting U(r)=0, this gives:

$$w^+(p) \cdot \left(q_a^j\right)^\alpha - w^-(1-p) \cdot \lambda \cdot \left(-q_b^j\right)^\beta = 0. \tag{6}$$

Solving for λ gives the following expression:

$$\lambda = \frac{w^+(p) \cdot \left(q_a^j\right)^\alpha}{w^-(1-p) \cdot \left(-q_b^j\right)^\beta}. \tag{7}$$

3. Experiment

3.1. Design

The experimental design consisted of two different treatments, with each respondent being allocated randomly to one of them. In both treatments it was made clear to the subjects that they should imagine having a particular QoL in terms of a percentage of full health. The QoL obviously cannot get higher than 100%, and, hence, 100% was the maximum of our outcome range. In order to prevent QoL from

getting close to 0, which may cause extreme behavior and subjects confusing QoL and life expectancy issues (Bleichrodt et al., 2003; Stiggelbout and de Haes, 2001), the minimum was set to 20%.

The first treatment (the Reference-Dependent Treatment: RDT) used two different reference points for gains and losses, and a mixed prospect. We used the minimum outcome, 20%, as the reference point in the gain part. Similarly, we took the maximum outcome, 100%, as the reference point in the loss part. Taking the minimum [maximum] as the reference point allowed for eliciting risk attitude for gains [losses] over the entire observed QoL range of 20%-100%. In fact, we used the same outcomes twice, but obtained from a different reference point. Therefore, we test the stability of risk attitude in gains and losses, and, hence, we perform a within-subject test of traditional EU, which predicts the same risk attitude for both reference points given that the final health states are the same. On the opposite, different risk attitudes are evidence in favor of different reference points. Finally, we set the reference point for the mixed prospect at the value exactly halfway this range, at 60%. The mixed prospect still allowed us to estimate a gain/loss asymmetry index (i.e., computing loss aversion while assuming linear utility and no probability weighting) and the amount of risk aversion in comparison to the amount of risk aversion in the gain and loss parts.

The second treatment (the Sign-Dependent Treatment: SDT) induced a unique reference point of 60% of full health, and proceeded by eliciting risk attitudes over both gains and losses as seen from this point. This yielded both gains and losses ranging between 0 and 40% (i.e. up to 100% and down to 20%). Since the reference point was the same for gains and losses in SDT, we could estimate a loss aversion index for this treatment. A mixed prospect was used for this purpose.

3.2. Main Experiment

3.2.1. Subjects and procedure

The experiment was conducted by a professional internet sampling company (Survey Sampling International). This company has much experience with internet surveys and a large representative database of subjects. A total of 500 subjects, representative for the Dutch general population, participated in the experiment. The subjects were rewarded with a monetary amount to be given to a charity fund of their choice. The first 37 respondents performed a pilot experiment to test whether everything worked correctly and whether it generated sensible responses. The remaining 463 respondents were included in the analysis. This sample consisted of n=227 in RDT and n=236 in SDT.

The experiment started with some questions regarding background characteristics. We gathered information about age, gender, number of children, marital status, income, education, health status (as classified by EQ-5D-5L-5L) and rating of health (according to a VAS). Subsequently, before starting the main experiment, elaborate instructions and practice questions were provided, explaining the kind of trade-offs that had to be made in the experiment. These questions also contained some

dominant options. If a subject, after repeated instruction, still chose the dominated option, we took this as a case of misunderstanding and screened these subjects out. Hence, the completed dataset only contains the results of subjects who picked the dominant options in the practice questions, strengthening the reliability of the data.

Indifferences were elicited by a combination of iterative choices and matching. We started with a bisection procedure that adjusted the value of $\bar{q}^j$ upwards or downwards depending on the chosen option. We used two of these choices to zoom in on an indifference point. In the first question, $\bar{q}^j$ was always equal to the expected value (EV). Having narrowed down the answer range after these two choices, we gave the residual range of possible indifference values, and asked the subject to express their indifference value by using a slider.

In case of indifference at the first choice of the iteration process, respondents were asked whether they were really indifferent. If they indicated not to be indifferent on second thought, they returned to the first choice and answered it anew. Otherwise, they would proceed to the next question, in which they were asked to explain their choice. At the end of the iteration they were asked to type in a value for the sure outcome such that they would be indifferent between the sure outcome and the risky prospect. If they were truly indifferent in the first choice, this value should have been equal to the value shown to them in the first choice (i.e., the EV). For respondents who were indifferent in the first choice of an iteration, we set $\bar{q}^j$ equal to the EV in the analysis (instead of using their answer to the open question).

3.2.2. Stimuli

The utility for both gains and losses was elicited by seven questions each. The order of the seven questions was random. In the gain part, the subjects were given the opportunity to choose one of two treatments that could improve their health status. The safe option contained a treatment that would increase QoL with $\bar{q}$ percentage points for certain, whereas the risky option involved a treatment that would gain two different percentage points, both with probability 0.5. We picked a probability of 0.5, because this prevented the task from being an excessive cognitive burden on subjects, and because it is the most common probability used in the literature on risky decision making.

All QoL values were contained in the interval [20%,100%]. We specified all QoL improvements to be transitional, lasting for 1 year. Subjects were told that afterwards their health would improve to (or remain at) full health. The reasons why we chose a 1 year duration were twofold. First, it was essential that the duration was the same for all respondents, in order to guarantee homogeneity of stimuli among respondents. If we took remaining lifetime, this would be different across respondents and, hence, this would add an additional source of unobserved heterogeneity and hamper comparability of the results. Second, a longer time span would decrease the survival

probability during this span. Respondents may then accordingly reduce their valuation of the health improvement, which would again impose additional unobserved heterogeneity among respondents.

The instructions of the gain part told the subjects to imagine that, because of a disease, their health status had deteriorated a while ago to a level of 20% of full health in RDT (60% in SDT). However, recently, the doctors had discovered two new treatments that could do something to combat the disease. These two treatments were the following. One treatment involved a sure gain. The other treatment was risky, giving a larger gain with probability 0.5, but a smaller gain (or none at all) with probability 0.5 as well. The amount of the gain in the riskless treatment was then elicited such that the respondent was indifferent between the two treatments. The experimental design created a distinction between past changes in health, to which the subject has already adapted, and likely changes. In the experimental instructions it was made clear that both changes in health were of a different nature to help the subject distinguishing the reference point (the current status quo) from its deviations (see Kahneman and Tversky, 1979, pp. 286-288 on this point).

In the loss part, the respondent's health status was about to deteriorate during the next year to a level of 20% of full health because of a disease. This part of the instructions was necessary to prevent inaction to be the optimal decision. In opposition to the deterioration to a level of 20% of full health in the gain part, the drop in QoL described in this part is only conditional on inaction and should not be considered as a possible status quo or reference point. After that year, the disease would disappear naturally, and their health would return to 100% in RDT (60% in SDT). However, two treatments were also available to reduce the loss encountered during the coming year. One treatment involved a sure loss. The other treatment was risky, giving a larger loss with probability 0.5, but a smaller loss (or none at all) with probability 0.5 as well. The amount of the loss in the riskless treatment was then elicited such that the respondent was indifferent between the two treatments. A translation of the full instructions is available in Appendix B. For mixed prospects, the subjects were instructed to assume that their health had deteriorated recently to a level of 60% of full health due to unknown causes. Without treatment, the disease would disappear after one year, causing their health to return to 100%. However, a treatment had recently become available that could improve their health immediately. The effects of this treatment were uncertain, because for half of the patients it generated serious side-effects, reducing their health even further, to a level lower than 60%. For the other half of the patients, the treatment had no side-effects and their health immediately improved to a level higher than 60%. The question was elicited in two different ways. First, we elicited q_b such that $(\frac{1}{2}, 20;\ -q_b) \sim 0$, that is, the subject could choose between no treatment and staying at QoL=60%, or taking a gamble with probability 0.5 of gaining 20 percentage points of health (from 60 to 80) and probability 0.5 of losing q_b percentage points of health (from 60 to 60-q_b). Second, we elicited q_a such that $(\frac{1}{2}, q_a, -20) \sim 0$, that is, the subject could choose between no treatment and staying at QoL=60%, or taking a gamble with probability 0.5 of losing 20 percentage points of health and a

probability of gaining q_a percentage points. Half of the subjects did the first variant and half did the second variant. They were allocated randomly to the different variants.

The experiment always started with the gain part, because we learnt from pilot sessions that this made it easier for subjects to understand the choice task. After that, the loss and mixed prospects were asked in random order. Li et al. (2015) and Vieider et al. (2015) also recommended to have gain questions preceding loss questions for this reason. Moreover, Etchart-Vincent and l'Haridon (2011) and Vieider et al. (2015) tested for order effects between gains and losses in this regard, and found none. Table 1 presents a list of all prospects used in the Main Experiment.

Table 1. List of prospects (in terms of final health levels)

RDT	Gains	Losses
1	(0.5,40;20)	(0.5,40;20)
2	(0.5,60;20)	(0.5,60;20)
3	(0.5,100;20)	(0.5,100;20)
4	(0.5,70;30)	(0.5,70;30)
5	(0.5,90;50)	(0.5,90;50)
6	(0.5,100;60)	(0.5,100;60)
7	(0.5,100;80)	(0.5,100;80)
SDT		
1	(0.5,70;60)	(0.5,60;50)
2	(0.5,85;60)	(0.5,60;35)
3	(0.5,100;60)	(0.5,60;20)
4	(0.5,90;70)	(0.5,50;30)
5	(0.5,90;90)	(0.5,40;30)
6	(0.5,100;80)	(0.5,40;20)
7	(0.5,100;90)	(0.5,30;20)

3.3. Robustness Experiment

In the Main Experiment, we observed a large amount of responses that implied risk neutrality. In order to test to what extent this has influenced our main findings, we ran a robustness experiment, in which we modified several aspects of the design. First, we removed the no preference option in the first questions, forcing subjects to choose one of the two options. If subjects were truly indifferent, this would not distort the elicitation, since they were still able to express the initial CE value as their indifference value at the end of the iteration. Second, we marginally changed the stimuli such that they did not consist of round numbers anymore. Using non-round stimuli made it harder to compute the EVs and respond accordingly with an EV heuristic (Bostic et al., 1990; Cohen et al., 1987; Pennings and Smidts, 2003; Wakker and Deneffe, 1996). Third, we framed the outcomes in terms of gains and losses instead of final health levels. Fourth, the slider was removed. Instead, we increased the number of choices and stopped after the CE range was narrow enough to have an indifference estimate that was of sufficient precision. This estimate was computed as the midpoint of the remaining range. Fifth, the CE provided in the first question was not necessarily equal to the EV anymore. The stimuli of the

Robustness Experiment, including the starting points of the sure options, are reported in Tables C2 and C3 in Appendix C. A total of 516 respondents participated in the Robustness Experiment, including n=253 in RDT and n=263 in SDT. The sample was again representative for the Dutch adult population in terms of age, gender, and education.

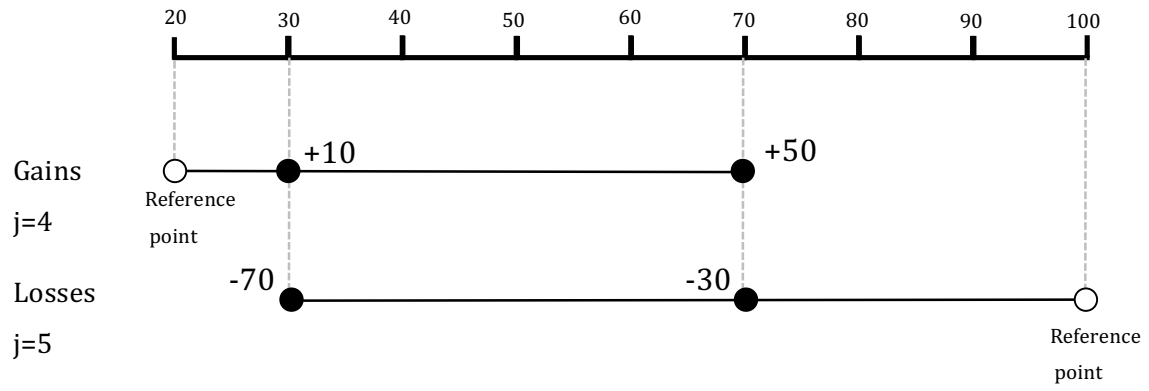
3.4. Analysis

3.4.1. Testing EU

Our design allowed for several tests of the descriptive validity of traditional EU in the domain of health-related QoL. First, within RDT we could compare risk attitudes for gains and losses in a setting similar to Kahneman and Tversky (1979) (problems 11 and 12). In this setting, if traditional EU holds, the reference points should not matter and, hence, both framings should return the same answers in terms of final health. The stimuli of the gain and loss parts of RDT were the same when expressed in final terms (see Figure 1 and Table 2). This made it possible to test EU in two ways:

1. Comparing the answers of the gain and of the loss part where the final outcomes are the same (j=1 for losses and j=7 for gains, and similarly for j=2 vs. 6, 3 vs. 3, 4 vs. 5, 5 vs. 4, 6 vs. 2, and 7 vs.1). Table 2 summarizes these comparisons.
2. Estimating the power utility parameter separately for the gain part and the loss part and check for consistency. These powers cannot be compared directly, because a value smaller [greater] than 1 reflects concavity [convexity] for gains, but convexity [concavity] for losses. Therefore, we also estimate a utility power on the CEs of the final health levels for gains and losses, where we hypothesize that utility is not smooth and, hence, not defined over final health.

Figure 1: Illustration of the equivalence of RPT prospects in final values for gain prospect (0.5,50;10) and loss prospect (0.5,-70;-30).



3.4.2. Elicitation of PT

The SDT allows for an elicitation of PT for QoL outcomes and to obtain a first measurement of its components. We do not have the exact same outcome stimuli in these two treatments, but since they cover the same QoL range of 20%-100%, we can compare the shapes of utility again. It may be the case that, if we find no difference between utility parameters for gains and utility for losses within RDT (giving no evidence against EU), we still find the combined utilities for gains and losses in SDT to differ from the elicited utility in RDT. Similar utilities in the two treatments would confirm EU, whilst a kink around 60% in SDT but not in RDT would be consistent with PT.[4]

We test this by comparing answers in SDT and RDT for similar prospects between subjects. For RDT, we use the large stake gambles $\left(\frac{1}{2}, 80, 0\right)$ for gains and $\left(\frac{1}{2}, 0, -80\right)$ for losses. Both gambles correspond to a gamble $\left(\frac{1}{2}, 100, 20\right)$ over final levels. For SDT we use the mixed gambles $\left(\frac{1}{2}, q_a, -20\right)$ and $\left(\frac{1}{2}, 20, -q_b\right)$. Both gambles correspond to medium stake gambles in terms of final wealth $\left(\frac{1}{2}, 60 + q_a, 40\right)$ and $\left(\frac{1}{2}, 80, 60 - q_b\right)$. The intuition is that within the EU framework, if there is risk aversion over modest stake gambles (in SDT), then at least the same amount of risk aversion is predicted over large stake gambles (in RDT) because of diminishing marginal utility. However, if it turns out that risk aversion is instead more important over medium stake gambles than over large stake gambles, this could be due to loss aversion affecting the mixed gamble in SDT, but not the large stake gambles in RDT.

Another test for loss aversion we perform, is to compare the amount of risk aversion found in the mixed prospect to that found in the gain and loss parts. In the RDT we have only one question for the reference point at 60% and we cannot perform any other tests of loss aversion but still it may give some indication of its presence or absence. That is, if there is no loss aversion, the starting level does not matter and, hence, the mixed prospect should give a similar amount of risk aversion as the other two parts. Instead, in case of loss aversion more risk aversion would be expected in the mixed prospect because respondents are less willing to choose the gamble[5].

If traditional EU is rejected, it is necessary to consider a more general decision model allowing for reference-dependence, like PT. Instead of identical curvature, the most common form of PT would predict convex utility for losses and concave utility gains; i.e., we would expect a lower

[4] At least if people take the induced RP as their status quo. It may also be true that people have a fixed RP of, say, 40%, which they use everywhere. Subjects would then reframe all outcomes as gains and losses as seen from 40% and the prediction would be a kink at 40% *in both treatments*.
[5] However, see Ert and Erev (2013) for the opposite finding.

answer (in terms of final values) to j=1 for gains than j=7 for losses, and likewise for the other comparisons in RDT.

The SDT allows for a more precise evaluation of sign-dependent decision models. For the SDT, we do not have the same outcomes in the gain and loss parts, because the initial health level was the same for both, and, hence, the final outcomes could by definition not be in the same range. Therefore, no direct comparison between the gain and loss parts was possible, but we could test for a gain/loss asymmetry in three ways:

1. Comparing the number of risk averse responses.
2. Investigating the power estimates for gains and losses.
3. A kink in SDT may be considered an indication of loss aversion. The presence of a kink at 60% in SDT is tested by estimating the loss aversion index: if the estimate of λ (Eq. 7) is higher than 1, this would be evidence of sign-dependence and loss aversion for outcomes below this reference point.

3.4.3. Exclusion criteria

In both treatments, it was not possible to estimate the parameters for some subjects who expressed indifference at the lowest or highest value of the prospect in multiple or all of the prospects implying a violation of stochastic dominance. These subjects were therefore removed from the parametric analysis (5 subjects in RDT and 9 subjects in SDT). In addition, no loss aversion index could be estimated for those subjects who were not prepared to accept any loss, as this would mean an infinite loss aversion index (30 [33] subjects in RDT [SDT]).

4. Results

4.1. Reliability

A minority of the subjects violated dominance in one or more questions, in that their indifference value was equal to the lowest or highest outcome of the prospect (RDT gains: lowest value 2.3%, highest 4.0%; RDT losses: lowest 2.9%, highest 2.4%; SDT gains: lowest 2.5%, highest 4.6%; SDT losses: lowest 2.8%, highest 6.9%).

Stochastic dominance, defined by the assignment of a higher indifference value to a prospect with at least one better outcome and no worse outcome, was also sometimes violated. In RDT, the

average rate of violation of FSD was equal to 2.16%. In SDT, the average rate of violation was equal to 2.56%.

4.2. CEs and risk attitude

A subject was classified as risk averse [risk seeking] if at least 5 out of 7 CE questions produced a risk averse [seeking] answer (i.e., a CE lower [higher] than the expected value of the prospect). This allowed taking into account response error. Because the data were not normally distributed, we performed nonparametric statistical tests (Wilcoxon signed ranks tests for within-subject analyses and Mann-Whitney tests for between-subjects analyses). Two-tailed p-values are reported.

4.2.1. RDT

Table 2 displays the quartiles of the CEs for each prospect of the Main Experiment. For gains and a reference point of 20, 35.9% [19.8%, 44.3%] of the answers was consistent with risk aversion [risk seeking, risk neutrality]. In the loss part, 42.9% [15.2%, 41.9%] of the answers reflected risk aversion [risk seeking, risk neutrality] for a reference point of 100. The within-subject correlation between the number of risk averse [seeking] responses for gains and the number of risk averse [seeking] responses for losses was high (RA: 0.67, RS: 0.58, $p<0.01$).

In the Robustness Experiment (see Appendix C for more details), we found more evidence of risk aversion (and almost no risk neutrality). For gains 58.4% [38.9%, 2.7%] of the answers was consistent with risk aversion [risk seeking, risk neutrality]. In the loss part, 62.0% [38.0%, 0%] of the answers reflected risk aversion [risk seeking, risk neutrality] for RP=100. In summary, it seems that the change in the design not only reduced the percentage of risk neutral subjects but also increased risk aversion. It seems that RN subjects of the Main Experiment split evenly between RA and RS.

Table 2. Median CEs in the Main Experiment and Wilcoxon tests on the difference between gains and losses in RDT

RDT	Gains	Prospect	Median (IQR)	Losses	Prospect	Median (IQR)	(p-value)
1		(0.5,40;20)	30 (30-30)		(0.5,40;20)	30 (28-30)	0.0049
2		(0.5,60;20)	40 (30-40)		(0.5,60;20)	40 (35.5-40)	0.5848
3		(0.5,100;20)	60 (40-60)		(0.5,100;20)	40 (40-60)	0.0001
4		(0.5,70;30)	50 (40-50)		(0.5,70;30)	50 (40-50)	0.0537
5		(0.5,90;50)	70 (60-70)		(0.5,90;50)	70 (60-70)	0.0597
6		(0.5,100;60)	80 (70-80)		(0.5,100;60)	74 (70-80)	0.0020
7		(0.5,100;80)	90 (85-90)		(0.5,100;80)	90 (85-90)	0.0048
SDT							
1		(0.5,70;60)	65 (65-66)		(0.5,60;50)	55 (53-55)	
2		(0.5,85;60)	72 (69-74)		(0.5,60;35)	47 (42-48)	
3		(0.5,100;60)	80 (70-80)		(0.5,60;20)	40 (30-40)	

4		(0.5,90;70)	80 (80-80)		(0.5,50;30)	40 (35-40)	
5		(0.5,90;90)	85 (85-85)		(0.5,40;30)	35 (35-35)	
6		(0.5,100;80)	90 (85-90)		(0.5,40;20)	30 (26-30)	
7		(0.5,100;90)	95 (92-95)		(0.5,30;20)	25 (25-25)	

Note: p-value corresponds to a Wilcoxon matched-paired test on the difference between certainty equivalents for the gain prospects and loss prospect when expressed in final health levels.

Comparison of gains and losses serves as a basis to test EU. When comparing gains to losses in RDT, we found more risk aversion for losses than gains ($p<0.01$), for 4 out of 7 CEs, and 5 out of 7 at the 6% level. Only for j=2 vs. j=6 there was no difference ($p=0.58$). For j=5 for gains vs. j=4 for losses the sign of the difference was the opposite: risk aversion was marginally lower for losses than for gains ($p<0.10$). Hence, we find a violation of EU in most comparisons. The final health levels were not always exactly the same in the gain part as in the loss part in the Robustness Experiment, so we did not perform this test there.

4.2.2. SDT

For gains, 33.7% [18.8%, 47. 5%] of the answers were consistent with risk aversion [risk seeking, risk neutrality]. In the loss part, this pattern was similar: 33.1% [17.1%, 49.8%] of the answers reflected risk aversion [risk seeking, risk neutrality]. Hence, fewer answers were consistent with risk aversion in SDT than in RDT. The within-subject correlation between the number of risk averse [seeking] responses for gains and the number of risk averse [seeking] responses for loss was again high (RA: 0.54, RS: 0.65; $p<0.01$). The classification of subjects in terms of risk attitude for gains and losses is similar.

Results from the Robustness Experiment were similar for SDT as for RDT. For gains, 58.0% [34.3%, 7.7%] of the answers was consistent with risk aversion [risk seeking, risk neutrality]. In the loss part, this pattern was similar: 57.6% [34.8%, 7.7%] of the answers reflected risk aversion [risk seeking, risk neutrality]. There was again more risk aversion than risk seeking in both domains ($p<0.01$).

4.2.3. Mixed prospects

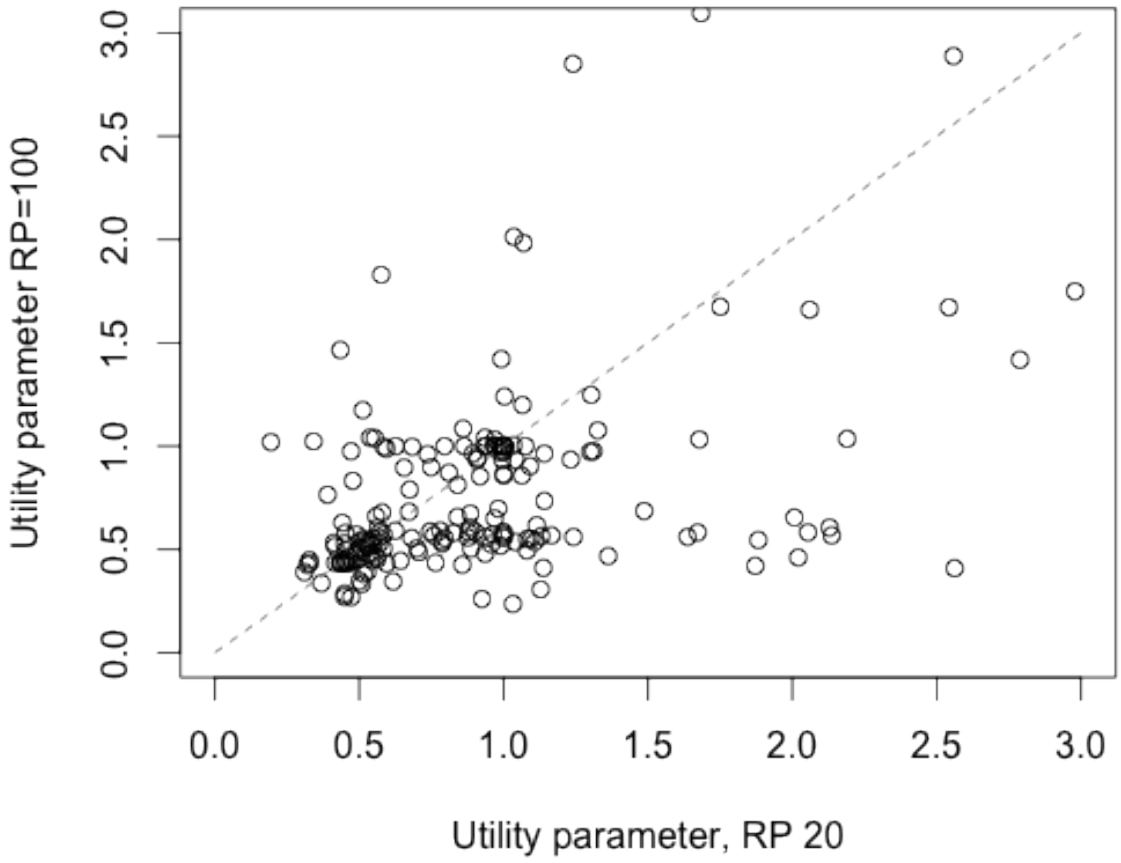
In the mixed prospect of the Main Experiment, 51.8% of the responses were risk averse, 23.8% were risk seeking, and 24.4% were classified as risk neutral. In the Robustness Experiment, 71.5% of the responses were risk averse and 28.5% were risk seeking. A more precise look at the results shows that this degree of risk aversion was comparable to the degree of risk aversion among non-risk neutral subjects in the Main Experiment (68.5%) and to the degree of risk aversion in previous studies eliciting loss aversion (Abdellaoui et al., 2008; Attema et al., 2013).

We observed more risk aversion in the mixed SDT prospect (53% in the Main Experiment and 66.9% in the Robustness Experiment) than in the large-scale RDT gain prospect (49.3% in the Main Experiment and 60.5% in the Robustness Experiment). Therefore, utility in SDT was not compatible with utility in RDT for gains, which may be due to loss aversion in the mixed SDT prospect. For losses, the test was inconclusive because we observed more risk aversion in the large-scale RDT prospect (63% in the Main Experiment and 75.5% in the Robustness Experiment) than in the mixed SDT prospect (53% in the Main Experiment and 66.9% in the Robustness Experiment), which could be the result of both diminishing marginal utility and loss aversion.

4.3. Tests using parameter estimates

Assuming EU, the power estimate of utility clearly differed between the loss and gain parts in RDT. Figure 2 shows the relationship between the utility parameter estimates under EU in the gain and loss parts. When the reference point was set to 20, the median utility parameter estimate was equal to 0.90. When the reference point was set to 100, for the same prospects over final health level, the median parameter estimate was lower and equal to 0.61 ($p<0.01$). Hence, utility was more concave when prospects were presented as losses than as gains. This result underlines the necessity to take into account sign-dependence (i.e., a different evaluation of outcomes depending on whether they are gains or losses) in decision models, PT or a sign-dependent version of EU.

*Figure 2. Comparison of parameter estimates under EU in RDT using the two different reference point levels RP=20 (gain part) and RP=100 (loss part)**



* We restricted the data points to cover the range of utility parameters [0-3]. Narrowing the range excluded 8.0% of the data points from the Figure.

Under PT, the power estimates for losses confirmed concavity (Tables 3 and 4; $p<0.01$ for both RDT and SDT, and for both experiments). In other words, respondents had concave utility for losses. No deviation from linearity was observed for utility for gains ($p>0.21$), except for RDT in the Robustness Experiment, where utility was concave ($\alpha=0.80$, $p<0.05$). Hence, we find no evidence of reflective utility, although we observe more utility curvature for losses than for gains.

We found loss aversion in both experiments (λ in SDT>1, $p<0.01$). The amount of loss aversion was robust across the two experiments. It shows that this is a robust phenomenon, occurring even in an experiment where we observe a lot of risk neutrality in the gain and loss prospects (i.e., the Main Experiment). The amount of loss aversion (a bit below 2) is somewhat lower than that frequently

observed for monetary outcomes (which tends to be a bit above 2, e.g. Tversky and Kahneman, 1992). In addition, the interquartile ranges of λ indicate a high amount of heterogeneity among individuals.

Risk aversion in RDT was captured by underweighting the probability of the best outcome (Main Exp.: median ω^+=0.46, p<0.01; Robustness Exp.: ω^+=0.41, p<0.01). The median estimate of ω^+ for SDT was exactly 0.50 in the Main Experiment, suggesting no systematic deviation from linearity in probabilities, but in the Robustness Experiment there was underweighting (0.35, p<0.01). For losses we found underweighting of the bad outcome in both treatments and in both experiments (p<0.01 for all four comparisons). Finally, no significant differences between the decision weights for gains and losses were found in the Main Experiment and RDT in the Robustness Experiment, but decision weights were lower for gains than for losses for SDT in the Robustness Experiment (p<0.01). The estimated decision weights are similar to previous observations of probability weighting of probability 0.5 using health outcomes (Bleichrodt and Pinto, 2002; Bleichodt et al., 1999).

Table 3a. Estimation results power model RDT, based on individual data (after removing 11 subjects choosing many dominated options), Main Exp.

	α (gains)	ω^+	β (losses)	ω^-
Median	1	0.46	1.65	0.48
IQR	0.65-1.34	0.32-0.53	1.04-2.83	0.34-0.55
p-value	0.49	<0.001	<0.001	<0.01

Note: p-value corresponds to the p-value of a Wilcoxon test for the difference with 1 (utility parameters) and 0.5 (probability weighting parameters). Number of observations: n=206 for gains and n=200 for losses.

Table 3b. Estimation results power model SDT, based on individual data, Main Exp.

	α (gains)	ω^+	β (losses)	ω^-	λ (excl. inf)	λ (incl. inf)
Median	0.98	0.50	1.16	0.50	1.35	1.97
IQR	0.69-1.18	0.35-0.54	0.98-1.98	0.37-0.58	0.63-4.05	0.76-11.6
p-value	0.22	0.003	<0.001	0.17	<0.001	<0.001

Note: p-value corresponds to the p-value of a Wilcoxon test for the difference with 1 (utility and loss aversion parameters) and 0.5 (probability weighting parameters). Number of observations: n=225 for gains, n=217 for losses, n=215 for loss aversion.

Table 4a. Estimation results power model RDT, based on individual data, Robustness Exp

	α (gains)	ω^+	β (losses)	ω^-
Median	0.80	0.41	1.80	0.47
IQR	0.40-1.37	0.24-0.61	0.95-3.03	0.28-0.66
p-value	0.04	<0.001	<0.001	0.13

Note: p-value corresponds to the p-value of a Wilcoxon test for the difference with 1 (utility parameters) and 0.5 (probability weighting parameters). Number of observations: n=201 for gains and n=216 for losses.

Table 4b. Estimation results power model SDT, based on individual data, Robustness Exp.

	α (gains)	ω^+	β (losses)	ω^-	λ
Median	0.95	0.35	1.06	0.47	1.88
IQR	0.74-1.46	0.20-0.53	0.72-2.13	0.30-0.63	0.50-11.55
p-value	0.12	<0.001	<0.001	0.13	<0.001

Note: p-value corresponds to the p-value of a Wilcoxon test for the difference with 1 (utility and loss aversion parameters) and 0.5 (probability weighting parameters). Number of observations: n=213 for gains, n=229 for losses, n=188 for loss aversion.

5. Discussion

Good descriptive models are required to correctly model health-related decisions and to debias health state valuations. The structure of our study, viz. the RDT and SDT treatments, allowed us to test such decision models. We analyzed individual decision making with regard to quality of life and found evidence for reference-dependence as well as loss aversion. However, the high amount of risk neutral people suggested an internal validity issue on separate domains. We ran a robustness experiment in order to remove this issue, and its results supported the results of the main experiment, although we also observed some notable differences. In particular, because of the use of different stimuli (making it less easy to use EV's as default options) and the removal of an indifference option, there was substantially less risk neutrality in the robustness experiment. This reduction resulted in a higher ratio of both risk averse and risk seeking respondents, with the ratio of risk averse to risk seeking respondents increasing slightly compared to the main experiment. This was translated into somewhat

more concave utility functions and more pronounced probability weighting in the robustness experiment.

We reported evidence of risk aversion for both gains and losses. It seems that the magnitude of the gains does not matter for the degree of risk aversion, but the sign of the outcomes as either gains or losses does, with more risk aversion present in the loss domain than in the gain domain. This risk aversion could be attributed to utility curvature as well as probabilistic pessimism.

Our findings of concave utility for losses confirm the results observed in the domain of life duration (Attema et al., 2013) and in pain intensity (Schosser et al., 2015). Therefore, this study reinforces the evidence that health is a fundamentally different commodity than money, with a positive correlation between concavity for gains and for losses. Instead, 'diminishing sensitivity' or reflection at the individual level, with a positive correlation between concavity of utility for gains and convexity for losses, is often found in the monetary domain (Abdellaoui et al., 2007; Abdellaoui et al., 2013; Budescu and Weiss, 1987; Schoemaker, 1990). However, the observed risk aversion in both gains and losses may also have been an artefact of the within-subject design of our study; that is, respondents could have had a preference to be consistent across domains, which would explain the high inter-domain correlation. This possibility may be tested in future research by adopting a between-subject design where each respondent only performs either a gains task or a loss task.

With money, convexity in the loss domain implies that people are more sensitive to losses near their status quo, than to the same losses remote from their status quo. For example, people may evaluate a loss of €11,000 as similar to a loss of €10,000, since the additional loss of €1,000 is not perceived as making much of a difference in the context of an already large loss. However, the difference between €1,000 and €2,000 is perceived as large. The difference is explained by a general characteristic of perception of quantities and numbers. To the contrary, the standard economic argument, diminishing marginal utility, reflects the intrinsic value of money and would predict concavity in the loss domain. In the health domain, with 60% as reference point, convexity would mean that a loss from 60% to 40% of QoL would be perceived as much "larger" than a loss from 40% to 20%. We observed the opposite in our experiments: the loss from 60% to 40% was perceived as smaller than the loss from 40% to 20%. A direct comparison with monetary outcomes implies that in the loss domain monetary outcomes are influenced by the aforementioned psychological perception of quantities, while this is not the case for health outcomes.

Our study also highlights that other elements of prospect theory can be transferred from the monetary domain to the health domain. Loss aversion appears to be a robust phenomenon, and is of a similar magnitude in this study as in estimations using monetary outcomes (Abdellaoui et al., 2007; Abdellaoui et al., 2008; Booij and van de Kuilen, 2009).

The findings in this study of concave utility for QALYs are not in line with the current application of QoL values. These applications usually assume a linear function over QALYs. Therefore, our results suggest that a modification to this habit may be appropriate. This result has

been observed under EU previously, but since EU is not descriptively valid, a wrong QALY model could have been confused with a wrong utility theory. Our result confirms that this is not the case and the assumption of a linear function over QALYs is not correct.

The experiments performed for this study have several limitations. First, QoL in numbers is perhaps hard to imagine, since people are more familiar with speaking about health in terms of ability to function in different activities, such as captured by the EQ-5D-5L classification system. However, in order to elicit utility, probability weighting and loss aversion a continuous scale is useful.

Second, since we did not know the respondents' reference points, we induced a status quo (i.e. an initial health state) that was the same for all respondents. Maybe respondents adopted the status quo as their reference points, but they may also have framed the outcomes as deviations from their own reference point (e.g., what they consider to be a normal health level for themselves), or they may have constructed a reference point according to some weighted average of these two levels. The latter would require more cognitive effort by the subjects, because they would then repeatedly have to change perspective from gains to losses, and vice versa, which would likely undermine data reliability. Moreover, our finding of differences between gains and losses in RDT suggests that people are sensitive to reference-dependence and changes in health rather than final health positions and take the induced starting value as their reference point. Furthermore, our finding of a difference in risk aversion between the mixed SDT prospect (with starting point 60%) and the large scale RDT loss prospect (with starting point 100%) suggests that subjects did not take 100% as their reference point in the mixed SDT prospect. In general, the lack of a theory about the formation of reference points when none are induced is one of the main weaknesses of prospect theory (Wakker, 2010). This is a problem in many contexts, not only experimental ones, and highlights the need for research about reference point formation (Arkes et al., 2008; Baucells et al., 2011; Kőszegi and Rabin, 2006).

Third, we could not separately estimate utility and the decision weight of probability 0.5, but had to estimate them simultaneously, with the risk of collinearities. Related to this, we measured only one decision weight, so that we have limited knowledge about the shape of the probability weighting function. This emphasizes the desire for a replication of our research that measures these concepts separately by means of two separate tasks, which would justify more robust conclusions.

Fourth, as has to be in health usually, our study did not use real (performance-based) incentives. Furthermore, we did not even pay subjects a flat fee, but instead let them pick a charity fund of their choice, to which our payment would be transferred. This decision was supported by an experiment with health outcomes by Bleichrodt and Pinto (2009), who found that most subjects declined to be paid for their participation.

Fifth, part of the respondents may have considered the stakes (temporary health shocks of 1 year) to be too small. This could explain the high amount of risk neutrality in the main experiment,

although the robustness experiment did not generate a lot of risk neutrality while using the same time horizon. Hence, future research using higher stakes is encouraged.

6. Conclusion

We can conclude that health is a special commodity with a number of behavioral differences as compared to, for instance, money. Nevertheless, several other phenomena are universal and occur in the health domain in a fashion similar as in the monetary domain. In particular, EU is violated for health outcomes as well as for monetary outcomes, with losses looming larger than gains and probabilistic pessimism. The most important difference seems to be a clear deviation from the commonly observed S-shaped utility (reflection); instead, universal concavity appears to be a robust pattern for health outcomes.

Appendix A. Derivation of predictions

In terms of the generalized QALY model, and assuming PT, the obtained indifferences for gains can be represented by:

$$V\left(\bar{q}^{j}\times FH_{i}\right)\times W(T) + V(FH_{i})\times[W(T) - W(1)] =$$

$$\omega^{+}\times V\left(q_{a}^{j}\times FH_{i}\right)\times W(1) + (1-\omega^{+})\times V\left(q_{b}^{j}\times FH_{i}\right)\times W(1) + V(FH_{i})\times[W(T) - W(1)]$$

Simplifying gives:

$$V\left(\bar{q}^{j}\times FH_{i}\right) = \omega^{+}\times V\left(q_{a}^{j}\times FH_{i}\right) + (1-\omega^{+})\times V\left(q_{b}^{j}\times FH_{i}\right)$$

Assuming a power function, $V\left(q^{j}\times FH_{i}\right) = \left(q^{j}\right)^{\alpha}\times {FH_{i}}^{\alpha}$, we obtain:

$$\left(\bar{q}^{j}\right)^{\alpha}\times {FH_{i}}^{\alpha} = \omega^{+}\times\left(q_{a}^{j}\right)^{\alpha}\times {FH_{i}}^{\alpha} + (1-\omega^{+})\times\left(q_{b}^{j}\right)^{\alpha}\times {FH_{i}}^{\alpha}$$

Simplifying and factoring out for ω^{+} yields:

$$\left(\bar{q}^{j}\right)^{\alpha} = \omega^{+}\times\left(\left(q_{a}^{j}\right)^{\alpha} - \left(q_{b}^{j}\right)^{\alpha}\right) + \left(q_{b}^{j}\right)^{\alpha}$$

Solving for $\bar{q}^{j}$ gives Eq. 4. For losses, the obtained indifferences yield similar equations:

$$V\left(\bar{q}^{j}\times FH_{i}\right) = \omega^{-}\times V\left(q_{a}^{j}\times FH_{i}\right) + (1-\omega^{+})\times V\left(q_{b}^{j}\times FH_{i}\right)$$

Assuming a power function, the utility after 1 year can also be canceled and we obtain:

$$-\left(-\bar{q}^{j}\right)^{\beta} = \omega^{-}\times -\left(-q_{a}^{j}\right)^{\beta} + (1-\omega^{-})\times -\left(-q_{b}^{j}\right)^{\beta}$$

And then:

$$\left(-\bar{q}^{j}\right)^{\beta} = \omega^{-}\times\left(\left(-q_{a}^{j}\right)^{\beta} - \left(-q_{b}^{j}\right)^{\beta}\right) + \left(-q_{b}^{j}\right)^{\beta}$$

In case of EU, we have w(p)=p and no reference-dependency. Hence, no distinction between gains and losses is made and we would predict no different results from the gain and loss parts. Therefore, EU gives the following indifference evaluation in both parts:

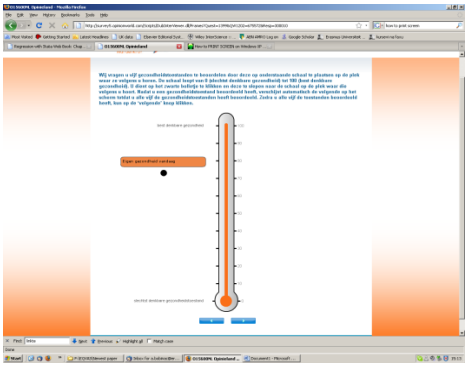
$$\left(\bar{q}^{j}\right)^{\alpha} = p\times\left(\left(q_{a}^{j}\right)^{\alpha} - \left(q_{b}^{j}\right)^{\alpha}\right) + \left(q_{b}^{j}\right)^{\alpha}$$

Appendix B. Translation of instructions

Gains

We are about to ask you questions about what you would do in case you were ill. Because these are difficult questions and given that you probably don't face this problem in reality, we will first ask you some simple questions which will hopefully help you in answering this questionnaire.

We will speak about health as a number. If we say that your health is 100, it means that your health is perfect. If we say that your health is equally bad as being dead, then we give your health a value of 0. Can you indicate how you would rate your own health today on a scale from 0 to 100?



In the following questions, assume that your health has deteriorated seriously a few years ago. The cause was unknown and doctors could not do anything to improve your health. Your health is, expressed on scale from 0 to 100, only 20 now, while it used to be 100 in the past.

How would you appreciate a health of 20 on a scale from 0 to 100?
-not very bad
-reasonably bad
-very bad
-extremely bad

Then now the good news. Doctors have found the cause of the problem. Not only that: two medical treatments are available. Both treatments ensure that the condition will completely resolve in one year, so that your health returns to the old level of 100. However, the treatments differ in the effect they

have during the coming year. We will ask you to choose between these two treatments, but first we will better explain the choice problem using a stepwise procedure.[6]

Example of a question

Question 1a.
The effects of Treatment A and B are indicated below. Which treatment would you choose in this case?

Your health without treatment: 20
(Remember that 100 means perfect health and 0 a health state that is equally bad as being dead)

Treatment A

	Health during next year	Health afterwards
Type I patient	30	100
Type II patient	30	100

Treatment B

	Health during next year	Health afterwards
Type I patient	40	100
Type II patient	20	100

❑ TREATMENT A ❑ BOTH TREATMENTS EQUALLY GOOD ❑ TREATMENT B

Losses

Practice question 1a.

Imagine you haven't felt very well the last time and go to the doctor. The doctor tells you that you have a disease that causes your health to deteriorate to 10 during the next year. After this year, the disease is self-limiting and your health will return to the original level of 100. Two treatments are available that can reduce the consequences of the disease.

Treatment A causes your health to drop no further than to 55.

[6] After these instructions, several practice questions followed. These are available upon request.

Treatment B causes full recovery in half of the patients with this disease (Type I patients): if you belong to this group, your health will stay at 100. The treatment doesn't work so well in the other half of the patients (Type II patients) and if you are in this group, your health immediately drops to 20.

Suppose you have to choose between the two treatments. It is not possible to determine in advance whether you are Type I or Type II. This will only be resolved once the treatment has started. Moreover, you cannot choose Treatment B first and then A, or vice versa. You can choose only one treatment.
Your initial health and the effects of the two treatments are summarized below. Can you indicate whether you would choose Treatment A or Treatment B?

Your health before onset of the disease: 100
(Remember that 100 means perfect health and 0 a health state that is equally bad as being dead)

Treatment A

	Health during next year	Health afterwards
Type I patient	55	100
Type II patient	55	100

Treatment B

	Health during next year	Health afterwards
Type I patient	100	100
Type II patient	20	100

❑ TREATMENT A ❑ BOTH TREATMENTS EQUALLY GOOD ❑ TREATMENT B

Mixed prospect

Practice question 1a.

In the following questions, assume that your health has deteriorated seriously a few years ago. The cause was unknown and doctors could not do anything to improve your health. Your health is, expressed on scale from 0 to 100, only 60 now, while it used to be 100 in the past.
After one year, the disease is self-limiting and your health will return to the original level of 100. However, doctors have recently developed a treatment that may do something about the disease, but the effects of this treatment are uncertain. In exactly half of the patients (Type I patients) the treatment works well. Those people immediately increase to a health level of 100. In the other half of the patients (Type II patients) the treatment causes many side-effects during the coming year. As a result, the health of those people immediately decreases to 40.

It is not possible to determine in advance who is Type I or Type II. This will only be resolved once the treatment has started.

The initial health of the group and effects of the two treatments are summarized below. Can you indicate whether you would choose the treatment or no treatment?

Your current health: 70

(Remember that 100 means perfect health and 0 a health state that is equally bad as being dead)

No treatment

	Health during next year	Health afterwards
Type I patient	60	100
Type II patient	60	100

Treatment

	Health during next year	Health afterwards
Type I patient	100	100
Type II patient	40	100

❑ NO TREATMENT ❑ NO PREFERENCE ❑ TREATMENT

APPENDIX C: Additional information on the stimuli

Table C1. Stimuli of the gain and loss prospects for SDT (in terms of both absolute values of changes and final values)

Absolute changes	j=1	j=2	j=3	j=4	j=5	j=6	j=7
$\lvert q_a \rvert$	10	25	40	30	30	40	40
$\lvert q_b \rvert$	0	0	0	10	20	20	30
Final levels Gains							
$60 + \lvert q_a \rvert$	70	85	100	90	90	100	100
$60 + \lvert q_b \rvert$	60	60	60	70	80	80	90
Comparison with RDT, losses	-	-	**	-	-	*	-
Final levels Losses							
$60 - \lvert q_a \rvert$	50	35	20	30	30	20	20
$60 - \lvert q_b \rvert$	60	60	60	50	40	40	30
Comparison with RDT, gains	-	-	**	-	-	*	-

Note: n.s.: difference not significant, *: significant at 10%, **: significant at 5%., -: not available.

Table C2. Stimuli of the gain and loss prospects for SDT (in terms of both absolute values of changes and final values) in the Robustness Experiment

Absolute changes	j=1	j=2	j=3	j=4	j=5	j=6	j=7
$\lvert q_a \rvert$	14	26	34	13	40	40	40
$\lvert q_b \rvert$	0	0	11	0	4	27	17
Final levels Gains							
$60 + \lvert q_a \rvert$	74	86	94	73	100	100	100
$60 + \lvert q_b \rvert$	60	60	71	60	64	87	77
St.point	66	72	84	67	81	95	87
Final levels Losses							
$60 - \lvert q_a \rvert$	46	34	26	47	20	20	20
$60 - \lvert q_b \rvert$	60	60	49	60	56	33	43
St.point	34	48	36	53	39	25	33

Table C3. Stimuli of the gain and loss prospects for RDT (in terms of both absolute values of changes and final values) in the Robustness Experiment

Absolute changes	j=1	j=2	j=3	j=4	j=5	j=6	j=7
$\lvert q_a \rvert$	23	38	80	47	70	78	80
$\lvert q_b \rvert$	0	2	4	10	33	40	57
Final levels Gains							
$20 + \lvert q_a \rvert$	43	58	100	67	90	98	100
$20 + \lvert q_b \rvert$	20	22	24	30	53	60	77
St.point	31	39	65	50	70	76	89
Final levels Losses							
$100 - \lvert q_a \rvert$	77	62	20	53	30	22	20
$100 - \lvert q_b \rvert$	100	98	96	90	67	60	43
St.point	89	81	55	70	50	44	31

APPENDIX D: Risk attitudes and observable characteristics.

In the experiment we gathered several observable characteristics of respondents. The following table shows the values of the observable characteristics of the subjects in both experiments for each RPT/SDT conditions.

Table D1: descriptive statistics, subject's characteristics

	Experiment 1				**Experiment 2**			
	RDT (n=227)		**SDT (n=236)**		**RDT (n=253)**		**SDT (n=263)**	
	Mean	**Std dev.**	**Mean**	**Std dev.**	**Mean**	**Std dev.**	**Mean**	**Std dev.**
Gender (1=Male)	0.49	0.50	0.47	0.50	0.50	0.50	0.51	0.50
Age	45.25	14.88	44.67	14.63	41.98	16.18	42.56	16.13
Education (from 1 to 7)	4.32	1.60	4.37	1.61	4.64	1.47	4.57	1.47
Income (from 1 to 13)	5	2.82	4.65	2.81	5.03	3.05	4.29	2.73
Children (0=none)	0.59	0.49	0.64	0.48	0.51	0.50	0.52	0.50
VAS	80.23	16.80	80.58	14.40	-	-	-	-
EQ-5D-5L	0.86	0.19	0.86	0.15	0.86	0.16	0.84	0.18

We used these background characteristics to investigate whether they could be related to the measurements of risk attitudes. The following tables show the results of ordinary least squares regression of the different components of risk attitudes (utility for gains and losses, probability weighting for gains and losses and loss aversion) on the background characteristics. Comparison of the results show that no systematic pattern of relation between risk attitudes and individual characteristics emerges in our study. For example, loss aversion increases with income in the main experiment, but the sign of the effect is opposite in the robustness experiment.

Table D2: Ordinary least square regression, results for RDT in the Main Experiment

	Dependent variable:			
	Utility parameter gains	probability weight gains	Utility parameter losses	probability weight losses
	(1)	(2)	(3)	(4)
Gender (female=1)	0.248	0.022	-0.466**	-0.012
	(0.188)	(0.025)	(0.228)	(0.025)
Age	0.013*	-0.001	-0.015*	0.001
	(0.007)	(0.001)	(0.008)	(0.001)
Education 1 (lower) to 7 (higher)	-0.118*	0.002	0.074	0.003
	(0.060)	(0.008)	(0.074)	(0.008)
Income 1 (lower) to 13 (higher)	-0.014	-0.001	0.039	0.002
	(0.033)	(0.005)	(0.041)	(0.005)
Children (none=0)	-0.074	0.014	0.007	0.006
	(0.214)	(0.029)	(0.261)	(0.029)
VAS	-0.002	0.001	-0.007	0.001
	(0.008)	(0.001)	(0.009)	(0.001)
EQ-5D-5L	0.021	0.052	0.575	-0.123
	(0.652)	(0.088)	(0.819)	(0.092)
Constant	1.350**	0.332***	2.541***	0.415***
	(0.656)	(0.089)	(0.840)	(0.094)
Observations	206	206	200	200

R^2	0.057	0.024	0.058	0.026
Adjusted R^2	0.024	-0.011	0.023	-0.010
Residual Std. Error	1.317 (df = 198)	0.178 (df = 198)	1.551 (df = 192)	0.173 (df = 192)
F Statistic	1.721 (df = 7; 198)	0.687 (df = 7; 198)	1.680 (df = 7; 192)	0.722 (df = 7; 192)

Note: *p<0.1; **p<0.05; ***p<0.01

Table D3: Ordinary least square regression, results for SDT in the Main Experiment

	Dependent variable:				
	Utility parameter gains	probability weight gains	Utility parameter losses	probability weight losses	loss aversion
	(1)	(2)	(3)	(4)	(5)
Gender (female=1)	-0.169	0.024	-0.005	-0.020	-0.197
	(0.227)	(0.022)	(0.276)	(0.025)	(2.032)
Age	0.003	-0.001	-0.001	0.0005	-0.010
	(0.009)	(0.001)	(0.011)	(0.001)	(0.080)
Education 1 (lower) to 7 (higher)	0.018	-0.004	0.123	-0.004	-0.665
	(0.074)	(0.007)	(0.090)	(0.008)	(0.680)
Income 1 (lower) to 13 (higher)	0.028	-0.006	0.010	-0.002	0.660*
	(0.042)	(0.004)	(0.051)	(0.005)	(0.369)
Children (none=0)	-0.013	-0.004	0.423	-0.026	-3.439
	(0.259)	(0.025)	(0.322)	(0.029)	(2.411)
VAS	0.021**	-0.001	0.020	-0.001	-0.080
	(0.010)	(0.001)	(0.012)	(0.001)	(0.092)
EQ-5D-5L	-2.349**	0.230**	-1.850	0.040	12.767
	(0.913)	(0.089)	(1.123)	(0.102)	(8.428)
Constant	1.365	0.433***	0.969	0.538***	2.493
	(0.861)	(0.084)	(1.055)	(0.096)	(7.829)
Observations	225	225	217	217	173
R^2	0.036	0.060	0.033	0.014	0.051

Adjusted R^2	0.005	0.029	0.001	-0.019	0.010
Residual Std. Error	1.575 (df = 217)	0.154 (df = 217)	1.875 (df = 209)	0.171 (df = 209)	12.374 (df = 165)
F Statistic	1.150 (df = 7; 217)	1.965* (df = 7; 217)	1.016 (df = 7; 209)	0.433 (df = 7; 209)	1.257 (df = 7; 165)

Note: *p<0.1; **p<0.05; ***p<0.01

Table D4: Ordinary least square regression, results for RDT in the Robustness Experiment

	Dependent variable:			
	Utility parameter gains	probability weight gains	Utility parameter losses	probability weight losses
	(1)	(2)	(3)	(4)
Gender (female=1)	-0.044	-0.032	0.168	-0.022
	(0.210)	(0.039)	(0.272)	(0.037)
Age	0.006	0.001	-0.002	0.0004
	(0.008)	(0.002)	(0.010)	(0.001)
Education 1 (lower) to 7 (higher)	0.106	0.014	0.036	-0.027**
	(0.075)	(0.014)	(0.097)	(0.013)
Income 1 (lower) to 13 (higher)	-0.040	-0.001	0.079*	-0.004
	(0.036)	(0.007)	(0.046)	(0.006)
Children (none=0)	0.208	-0.063	-0.324	0.010
	(0.270)	(0.050)	(0.344)	(0.047)
EQ-5D-5L	0.053	-0.020	0.030	-0.038
	(0.687)	(0.126)	(0.927)	(0.126)
Constant	0.428	0.392***	1.751	0.645***
	(0.784)	(0.144)	(1.083)	(0.147)
Observations	201	201	216	216
R^2	0.019	0.022	0.023	0.032
Adjusted R^2	-0.012	-0.009	-0.005	0.004
Residual Std. Error	1.452 (df = 194)	0.266 (df = 194)	1.945 (df = 209)	0.265 (df = 209)
F Statistic	0.610 (df = 6; 194)	0.716 (df = 6; 194)	0.819 (df = 6; 209)	1.142 (df = 6; 209)

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table D5: Ordinary least square regression, results for SDT in the Robustness Experiment

	Dependent variable:				
	Utility parameter gains	probability weight gains	Utility parameter losses	probability weight losses	loss aversion
	(1)	(2)	(3)	(4)	(5)
Gender (female=1)	-0.347	0.021	-0.032	-0.027	-0.311
	(0.213)	(0.035)	(0.214)	(0.036)	(2.656)
Age	-0.006	0.001	-0.012	0.001	0.046
	(0.008)	(0.001)	(0.007)	(0.001)	(0.090)
Education 1 (lower) to 7 (higher)	0.114	-0.019	0.032	-0.003	1.253
	(0.078)	(0.013)	(0.076)	(0.013)	(0.952)
Income 1 (lower) to 13 (higher)	0.032	0.002	-0.025	0.002	-0.956*
	(0.040)	(0.007)	(0.042)	(0.007)	(0.503)
Children (none=0)	0.232	-0.005	0.355	0.00001	-0.047
	(0.247)	(0.041)	(0.246)	(0.041)	(3.023)
EQ-5D-5L	-0.816	0.085	0.175	-0.027	1.018
	(0.627)	(0.104)	(0.642)	(0.107)	(8.030)
Constant	1.737**	0.362***	1.879**	0.484***	3.069
	(0.737)	(0.123)	(0.746)	(0.125)	(9.014)
Observations	213	213	229	229	163
R^2	0.048	0.018	0.017	0.007	0.028
Adjusted R^2	0.020	-0.011	-0.009	-0.020	-0.009
Residual Std. Error	1.532 (df = 206)	0.255 (df = 206)	1.590 (df = 222)	0.266 (df = 222)	16.499 (df = 156)

F Statistic	1.734 (df = 6; 206)	0.623 (df = 6; 206)	0.655 (df = 6; 222)	0.258 (df = 6; 222)	0.757 (df = 6; 156)

Note: *p<0.1; **p<0.05; ***p<0.01

References

-Abdellaoui M. Parameter-Free Elicitation of Utility and Probability Weighting Functions. Management Science 2000;46; 1497-1512.

-Abdellaoui M, Bleichrodt H, l'Haridon O. Sign-dependence in intertemporal choice. Journal of Risk and Uncertainty 2013;47; 225-253.

-Abdellaoui M, Bleichrodt H, l'Haridon O. A tractable method to measure utility and loss aversion under prospect theory. Journal of Risk and Uncertainty 2008;36; 245-266.

-Abdellaoui M, Bleichrodt H, Paraschiv C. Loss Aversion Under Prospect Theory: A Parameter-Free Measurement. Management Science 2007;53; 1659-1674.

-Abdellaoui M, l'Haridon O, Paraschiv C. Experienced vs. Described Uncertainty: Do We Need Two Prospect Theory Specifications? Management Science 2011;57; 1879-1895.

-Abellán-Perpinán JM, Bleichrodt H, Pinto-Prades JL. The predictive validity of prospect theory versus expected utility in health utility measurement. Journal of Health Economics 2009;28; 1039-1047.

-Arkes HR, Hirshleifer D, Jiang D, Lim S. Reference point adaptation: Tests in the domain of security trading. Organizational Behavior and Human Decision Processes 2008;105; 67-81.

-Attema AE, Brouwer WBF, l'Haridon O. Prospect theory in the health domain: A quantitative assessment. Journal of Health Economics 2013;32; 1057-1065.

-Baker R, Bateman I, Donaldson C, Jones-Lee M, Lancsar E. Weighting and valuing quality-adjusted life-years using stated preference methods: preliminary results from the Social Value of a QALY Project. Health Technology Assessment 2010;14; 161.

-Baucells M, Weber M, Welfens F. Reference-Point Formation and Updating. Management Science 2011;57; 506-519.

-Bleichrodt H, Abellan-Perpiñan JM, Pinto-Prades JL, Mendez-Martinez I. Resolving Inconsistencies in Utility Measurement Under Risk: Tests of Generalizations of Expected Utility. Management Science 2007;53; 469-482.

-Bleichrodt H, Miyamoto J. A Characterization of Quality-Adjusted Life-Years Under Cumulative Prospect Theory. Mathematics of Operations Research 2003;28; 181-193.

-Bleichrodt H, Pinto JL. A Parameter-Free Elicitation of the Probability Weighting Function in Medical Decision Analysis. Management Science 2000;46; 1485-1496.

-Bleichrodt H, Pinto JL. Loss aversion and scale compatibility in two-attribute trade-offs. Journal of Mathematical Psychology 2002;46; 315-337.

-Bleichrodt H, Pinto JL. New evidence of preference reversals in health utility measurement. Health Economics 2009;18; 713-726.

-Bleichrodt H, Pinto JL, Abellán-Perpinán JM. A consistency test of the time trade-off. Journal of Health Economics 2003;22; 1037-1052.

-Bleichrodt H, Pinto JL, Wakker PP. Making Descriptive Use of Prospect Theory to Improve the Prescriptive Use of Expected Utility. Management Science 2001;47; 1498-1514.

-Bleichrodt H, Quiggin J. Characterizing QALYs under a General Rank Dependent Utility Model. Journal of Risk and Uncertainty 1997;15; 151-165.

-Bleichrodt H, Schmidt U, Zank H. Additive Utility in Prospect Theory. Management Science 2009;55; 863-873.

-Bleichrodt H, van Rijn J, Johannesson M. Probability weighting and utility curvature in QALY-based decision making. Journal of Mathematical Psychology 1999;43; 238-260.

-Booij AS, van de Kuilen G. A parameter-free analysis of the utility of money for the general population under prospect theory. Journal of Economic Psychology 2009;30; 651-666.

-Bostic R, Herrnstein RJ, Luce RD. The effect on the preference-reversal phenomenon of using choice indifferences. Journal of Economic Behavior & Organization 1990;13; 193-212.

-Bruhin A, Fehr-Duda H, Epper T. Risk and Rationality: Uncovering Heterogeneity in Probability Distortion. Econometrica 2010;78; 1375-1412.

-Budescu DV, Weiss W. Reflection of transitive and intransitive preferences: A test of prospect theory. Organizational Behavior and Human Decision Processes 1987;39; 184-202.

-Cohen MD, Jaffray J, Said T. Experimental comparison of individual behavior under risk and under uncertainty for gains and for losses. Organizational Behavior and Human Decision Processes 1987;39; 1-22.

-Dolan P, Tsuchiya A. The social welfare function and individual responsibility: Some theoretical issues and empirical evidence. Journal of Health Economics 2009;28; 210-220.

-Ert E, Erev I. On the descriptive value of loss aversion in decisions under risk: Six clarifications. Judgment and Decision Making 2013;8; 214-235.

-Etchart-Vincent N, l'Haridon O. Monetary incentives in the loss domain and behavior toward risk: An experimental comparison of three reward schemes including real losses. Journal of Risk and Uncertainty 2011;42; 61-83.

-Furlong W, Feeny DH, Torrance GW, Barr RD. The Health Utilities Index (HUI®) system for assessing health-related quality of life in clinical studies. Annals of Medicine 2001;33; 375-384.

-Gächter S, Johnson E, Herrmann A. Individual-level loss aversion in riskless and risky choices. 2007. CeDEx Discussion Paper 2007-02.

-Kahneman D, Tversky A. Prospect Theory: An Analysis of Decision under Risk. Econometrica 1979;47; 263-291.

-Köbberling V, Wakker PP. An index of loss aversion. Journal of Economic Theory 2005;122; 119-131.

-Kőszegi B, Rabin M. A Model of Reference-Dependent Preferences. Quarterly Journal of Economics 2006;121; 1133-1165.

-Krischer JP. An Annotated Bibliography of Decision Analytic Applications to Health Care. Operations Research 1980;28; 97-113.

-Li Z, Rohde KIM, Wakker PP. Improving one's choices by putting oneself in others' shoes — an experimental analysis. Working paper. 2015. http://people.few.eur.nl/wakker/pdf/choicepredict.pdf.

-Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, Ciampi A, Till JE, Boyd NF. The measurement of patients' values in medicine. Medical Decision Making 1982;2; 449-462.

-Miyamoto JM, Eraker SA. A multiplicative model of the utility of survival duration and health quality. Journal of Experimental Psychology: General 1988;117; 3-20.

-Miyamoto JM, Eraker SA. Parametric models of the utility of survival duration: Tests of axioms in a generic utility framework. Organizational Behavior and Human Decision Processes 1989;44; 166-202.

-Pennings JME, Smidts A. The Shape of Utility Functions and Organizational Behavior. Management Science 2003;49; 1251-1263.

-Peters H. A preference foundation for constant loss aversion. Journal of Mathematical Economics 2012;48; 21-25.

-Pinto-Prades J, Sánchez-Martínez F, Corbacho B, Baker R. Valuing QALYs at the end of life. Social Science & Medicine 2014;113; 5-14.

-Pliskin JS, Shepard D, Weinstein MC. Utility functions for life years and health status. Operations Research 1980;28; 206-224.

-Riddel M, Kolstoe S. Heterogeneity in life-duration preferences: Are risky recreationists really more risk loving? Journal of Risk and Uncertainty 2013;46; 191-213.

-Schmidt U. Reference dependence in cumulative prospect theory. Journal of Mathematical Psychology 2003;47; 122-131.

-Schoemaker PJH. Are Risk-Attitudes Related Across Domains and Response Modes? Management Science 1990;36; 1451-1463.

Schosser S, Trarbach, JN, Vogt B. How does the perception of pain determine the selection between different treatments? Experimental evidence for convex utility functions over pain duration and concave utility over pain intensity. Journal of Economic Behavior & Organization 2015; in press.

-Stiggelbout AM, de Haes JCJM. Patient Preference for Cancer Therapy: An Overview of Measurement Approaches. Journal of Clinical Oncology 2001;19; 220-230.

-Tom SM, Fox CR, Trepel C, Poldrack RA. The Neural Basis of Loss Aversion in Decision-Making Under Risk. Science 2007;315; 515-518.

-Treadwell JR, Lenert LA. Health Values and Prospect Theory. Medical Decision Making 1999;19; 344-352.

-Tversky A, Kahneman D. Advances in Prospect Theory: Cumulative Representation of Uncertainty. Journal of Risk and Uncertainty 1992;5; 297-323.

-Vieider FM, Chmura T, Martinsson P. Risk attitudes, development, and growth macroeconomic evidence from experiments in 30 countries. Working paper. 2015. http://www.ferdinandvieider.com/risk_development_growth.pdf.

-Wakker PP. Explaining the characteristics of the power (CRRA) utility family. Health Economics 2008;17; 1329-1344.

-Wakker PP. Prospect theory: For risk and ambiguity, Cambridge: Cambridge University Press; 2010.

-Wakker P, Deneffe D. Eliciting von Neumann-Morgenstern Utilities When Probabilities Are Distorted or Unknown. Management Science 1996;42; 1131-1150.