Multivariate risk preferences in the QALY model

Attema, Arthur and Frasch, Jona and L'Haridon, Olivier

Erasmus University Rotterdam, Erasmus University Rotterdam, University of Rennes 1

9 July 2020

Multivariate risk preferences in the QALY model

ARTHUR E. ATTEMA
Erasmus School of Health Policy & Management, Erasmus University Rotterdam, attema@eshpm.eur.nl

JONA J. FRASCH
Erasmus School of Health Policy & Management, Erasmus University Rotterdam

OLIVIER L'HARIDON
University of Rennes, France
olivier.lharidon@univ-rennes1.fr

Abstract

In recent years the interest in multivariate and higher-order risk preferences has increased noticeably. A growing body of literature has demonstrated both the relevance and the impact of these preferences in several domains, although for health the empirical evidence is lacking. In this study we empirically measure multivariate and higher-order risk preferences for quality of life and longevity, the two elements of the Quality-Adjusted Life Year (QALY) model. We observe overwhelming support for correlation seeking between these two attributes as well as significant evidence of cross-imprudence and cross-intemperance. These findings indicate that higher-order risk preferences appear to deviate more from neutrality for health than for money. Furthermore, we test if preferences for a risky treatment for a disease affecting only quality of life, depend on life expectancy. Our results show no systematic evidence of such a relation, although there is a marginally significant positive relation between riskiness of the comorbidity affecting life expectancy and risk aversion for a treatment affecting quality of life. We therefore observe no definitive deviation from the QALY model, although the model appears to be more robust when expected longevity is high.

JEL Classification: I10

Key words: correlation attitude, prudence, QALYs, risk apportionment, risk aversion

1. Introduction

Health and health care are surrounded by a lot of risk, implying that risk aversion plays a central role in health economics. Recently, several studies have convincingly shown that also some concepts beyond risk aversion, such as prudence (i.e. a positive sign of the third derivative), are much more important than previously thought (e.g. Eeckhoudt and Schlesinger, 2006). These concepts are coined *higher-order risk attitudes*. Hence, the necessity to look beyond second-order risk attitudes has become clear, also in the health care field. This knowledge is important for several reasons. First, it allows to test if the quality-adjusted life year (QALY) model represents individual health preferences, and hence if QALYs are a proper metric to value health improvements. Second, higher-order risk preferences are relevant to many everyday health care decisions, such as risky treatment choices to combat a disease in the face of comorbidities. It is well known that many people suffer from two or more diseases at the same time (MacMahon, 2018), which may influence their preferences for treating their primary disease.

Courbage and Rey (2006) pointed out that the level of prudence is a main determinant of the optimal level of prevention for health risks, and Pauker (2014) advocated higher-order risk attitudes as a research topic that should receive priority on the research agenda in the domain of medical decision making. Moreover, Bleichrodt et al. (2003a) have shown the importance of higher-order risk attitudes in treatment decisions in the presence of comorbidities influencing life expectancy. They demonstrated that economic evaluations and medical decision analyses that ignore comorbidities will lead to recommendations that are biased in the direction of too much treatment if aversion to health status risks increases with life expectancy. They also derived several predictions regarding treatment decisions under particular assumptions, but so far these predictions had not yet been tested empirically. In addition, Eeckhoudt et al. (2007) showed how investment in tertiary preventive care (i.e., the treatment of an established or chronic disease in order to minimize the negative health consequences of the disease) depends on cross-prudence of health and income, i.e. it depends on whether an individual has a positive third cross-derivative of income with respect to health.

Krieger and Mayrhofer (2012) have explored higher-order risk attitudes in a health context empirically and observed both risk aversion and prudence. However, they only studied univariate risk attitudes and no multivariate risk attitudes, whereas in many settings a decision maker actually faces more than one attribute (Keeney and Raiffa, 1993). Eeckhoudt et al. (2007) and Ebert and van de Kuilen (2015) have stressed the importance of multi-attribute decision making, given the high prevalence of decisions where more than one attribute is involved. In the health domain, for instance, the widely used QALY model, which is the recommended metric to be used in health economic evaluations (Sanders et al., 2016), involves the attributes longevity and quality of life (QoL).

In case of two attributes, correlation aversion means that an individual prefers a 50% chance of a loss in one attribute and a 50% chance of a loss in the other attribute over a 50-50 gamble offering a loss in neither attribute or a loss in both (Eeckhoudt et al., 2007). An example of correlation aversion in health is when a patient prefers a lottery where he will get either a lower quality of life (50% chance) or a shorter life expectancy (50% chance) over a lottery where he has a 50% chance to get both a health deterioration and a lower life expectancy at the same time, and 50% chance to get no health losses at all. Bleichrodt et al. (2003a) showed that various consequences of the QALY model can be tested by obtaining knowledge about higher-order (cross-) derivatives of the utility function for longevity and QoL. One of their predictions was that people are risk averse for both longevity and quality of life, and correlation seeking for the combination of these two attributes. That is, people would prefer to combine a bad [good] health state with a short [long] life duration over mixing these two.[1] The risk apportionment technique allows us to test these predictions.

Bleichrodt et al. (2003a) also showed that, according to the QALY model, risk aversion for QoL should not depend on having a comorbidity that only affects longevity. In addition, they predicted that decreases in the riskiness of longevity caused by this comorbidity will generally lead to more treatment-prone behaviour (i.e. people get less risk averse for QoL). Finally, Bleichrodt et al. (2003a) derived how risk aversion, and

[1] This prediction was based on empirical evidence by McNeil et al. (1981) that people were not willing to trade off time to gain health for short life durations, and by Sutherland et al. (1982) that extension of lifetime beyond a certain threshold ('maximal endurable time') is valued negatively for poor health states.

hence treatment intensity, depend on higher-order multivariate risk preferences (i.e. risk aversion, correlation aversion, cross-prudence, and cross-temperance).

Attema et al. (2019) recently applied the risk apportionment technique to the health field, when they measured multivariate risk preferences, up to the fourth order, for longevity and wealth. They reported substantial risk aversion and correlation aversion for gains, but the opposite was found for losses. Furthermore, they observed less substantial amounts of prudence and temperance, but still significantly more than 50%. However, that study only investigated the duration component of the QALY model and hence could not test all the propositions from Bleichrodt et al. (2003a).

In this paper we are the first to empirically study several higher-order properties of the QALY model. This design enables us to test the theoretical predictions put forward by Bleichrodt et al. (2003a). In a nutshell, we combine an implementation of the risk apportionment technique with a treatment intensity task, in which we measure risk aversion for QoL for different life durations. First, we obtain evidence on individuals' correlation attitude between longevity and QoL. Second, we elicit their third- and fourth-order multivariate risk attitudes, i.e. cross-prudence and cross-temperance. Finally, we measure preferred treatment intensity for treating a disease affecting only QoL for patients also suffering from a comorbidity which affects longevity. Here, a higher treatment intensity increases the spread in the potential QoL outcomes. The latter measure enables us to test several theoretical predictions based on the QALY model as suggested by Bleichrodt et al. (2003a).

Our results show that subjects have marked risk preferences for longevity and QoL. First, we find a lot of risk aversion for both attributes, confirming most theoretical models. Second, we confirm Bleichrodt et al.'s (2003a) prediction of correlation seeking, with an overwhelming majority of subjects showing this preference. Furthermore, in contrast to most studies using monetary outcomes, we also find highly significant evidence for cross-imprudence and cross-intemperance. However, we observe no systematic correlation between treatment intensity and duration. Finally, we observe a marginally significant relation between treatment intensity and riskiness of life duration, in agreement with the intuition of Bleichrodt et al. (2003a).

2. Method

We assume preferences $\succcurlyeq$ satisfy a weak-order, i.e. they are complete and transitive. Individuals care about QoL (q) and longevity (t). According to the QALY model, preferences for chronic health states are evaluated by:

$$U(q,t) = V(q) \times W(t). \quad (1)$$

If expected utility holds, a subject is risk averse for QoL if $U_{qq} \leq 0$ and risk averse for longevity if $U_{tt} \leq 0$. Prudence for QoL holds if $U_{qqq} \geq 0$, prudence for longevity implies $U_{ttt} \geq 0$ and temperance holds if $U_{qqqq} \leq 0$ for QoL and $U_{tttt} \leq 0$ for longevity. Concerning multivariate risk preferences, a subject is correlation averse if $U_{qt} \leq 0$, cross-prudent for longevity if $U_{qqt} \geq 0$, cross-prudent for QoL if $U_{qtt} \geq 0$, and cross-temperate if $U_{qqtt} \leq 0$. Opposite signs define correlation seeking, cross-imprudence and cross-intemperance, respectively.

Eeckhoudt and Schlesinger (2006) were the first to operationalize (higher-order) risk preferences in terms of choices between two binary lotteries with equally likely outcomes that distribute harm and benefits differently, as illustrated below. An example of an item revealing risk aversion for QoL is the following:

What is your most preferred alternative?

Option A	*Option B*
50%: Live with 40% of full health for 40 years *50%: Live with 50% of full health for 40 years*	*50%: Live with 30% of full health for 40 years* *50%: Live with 60% of full health for 40 years*

Here, the risk averse individual would choose Option A, because it offers the same expected QoL as Option B (i.e. 45%), but with a lower spread. In fact, Option B is a mean-preserving spread of Option A. the general idea of the risk apportionment method is to have these kinds of choices between two-outcome gambles, with one resulting

from the other from a mean-preserving spread. Similarly, risk aversion for longevity could be determined by gambles such as the following:

What is your most preferred alternative?

Option A	*Option B*
50%: Live with 60% of full health for 40 years *50%: Live with 60% of full health for 40 years*	*50%: Live with 60% of full health for 30 years* *50%: Live with 60% of full health for 50 years*

In this example, Option A is riskless and Option B involves a mean-preserving spread of the same longevity. The risk apportionment method also allows for eliciting higher-order risk attitudes by adding different sources of uncertainty. For example, prudence for longevity can be elicited by the following choice:

What is your most preferred alternative?

Option A	*Option B*
50%: Live with 60% of full health for 40 years *50%: Live with 60% of full health for 10 OR 30 years*	*50%: Live with 60% of full health for 30 years OR 50 years* *50%: Live with 60% of full health for 20 years*

In this case, QoL is always 60% and longevity is either 40 years or 20 years. The choice involves distributing a zero-mean longevity risk of $\tilde{t} = \pm 10$ years to the bad longevity outcome (20 years, Option A) or the good longevity outcome (40 years, Option B). The former choice reflects *imprudence* and the latter choice reflects *prudence*. Similarly, temperance can be elicited by including two independent longevity or QoL risks and determining if the respondent prefers to aggregate (intemperance) or disaggregate (temperance) these risks.

Eeckhoudt et al. (2007) have demonstrated that the risk apportionment method can also be extended to elicit (higher-order) cross-risk attitudes when risk in both attributes is involved. For example, consider the following gamble:

What is your most preferred alternative?

Option A	*Option B*
50%: Live with 60% of full health for 40 years *50%: Live with 30% of full health for 20 years*	*50%: Live with 60% of full health for 20 years* *50%: Live with 30% of full health for 40 years*

This gamble involves risk in both QoL (30% or 60%) and longevity (20 or 40 years). The essential choice is if one prefers to combine the good outcome for QoL with the good outcome for longevity, while at the same time combining the bad outcomes for both (Option A), or if one prefers to spread the risks and combine the good outcome for the one attribute with the bad outcome for the other attribute (Option B). The former is deemed *correlation seeking* and the latter *correlation aversion*.
Tests of cross-prudence and cross-temperance can be conducted in a similar fashion. The below question could for instance be used for cross-prudence for longevity.

What is your most preferred alternative?

Option A	*Option B*
50%: Live with 60% of full health for 30 years *50%: Live with 40% OR 80% of full health for 40 years*	*50%: Live with 40% OR 80% of full health for 30 years* *50%: Live with 60% of full health for 40 years*

Looking closely, we can see that one lives either 30 or 40 more years in both gambles. Furthermore, QoL may be 60% or it may be another gamble, resulting in either 40% or 80%. In effect, a zero-mean risk on QoL ($\tilde{q}\sim\pm 20\%$) has to be apportioned to either the good outcome of the gamble (i.e. t=40 years, Option A) or the bad outcome of the

gamble (i.e. t=30 years, Option B). Someone who prefers to combine the zero-mean risk with the good longevity outcome is said to be cross-prudent for longevity, whilst someone who prefers combining the zero-mean risk with the bad longevity outcome is called cross-imprudent for longevity. Tests for cross-prudence for QoL and temperance can be done similarly.

Embedded in our study is the assumption that, generally, individuals prefer both higher levels of longevity and higher levels of QoL. While this method relies on the assumption that individuals aim to maximize their utility, it does not require assumptions about the functional form of the utility function (Attema et al., 2019). The risk apportionment technique can also be applied to elicit the other traits mentioned above.

In order to test the other predictions of Bleichrodt et al. (2003a), as described in the introduction, we elicit the sign of several (higher-order) risk traits. Table 1 gives an overview of all traits we elicited and the associated implications for the utility function in case of EU.

Table 1. Overview of elicited traits and their implied EU condition.

Trait if Prospect 1 is chosen	Prospect 1	Prospect 2	EU condition Prospect 1 is chosen
Risk aversion for QoL ($q > q_2 > q_1$)	$(0.5, q-q_1; q-q_2)$	$(0.5, q-q_1-q_2; q)$	$U_{qq} \leq 0$
Risk aversion for longevity (t> $t_2 > t_1$)	$(0.5, t-t_1; t-t_2)$	$(0.5, t-t_1-t_2; t)$	$U_{tt} \leq 0$
Correlation aversion ($q > q_1, t > t_1$)	$(0.5, t-t_1, q-q_1; q, t)$	$(0.5, t, q-q_1; q, t-t_1)$	$U_{qt} \leq 0$
Cross-prudence for QoL ($q > q_1, E(\tilde{t}) = 0$)	$(0.5, t, q-q_1; q, t+\tilde{t})$	$(0.5, t+\tilde{t}, q-q_1; t, q)$	$U_{qtt} \geq 0$
Cross-prudence for longevity (t> $t_1, E(\tilde{q}) = 0$)	$(0.5, t-t_1, q; q+\tilde{q}, t)$	$(0.5, t-t_1, q+\tilde{q}; q, t)$	$U_{qqt} \geq 0$
Cross-temperance ($E(\tilde{t}) = 0, E(\tilde{q}) = 0$)	$(0.5, t+\tilde{t}, q; q+\tilde{q}, t)$	$(0.5, t+\tilde{t}, q+\tilde{q}; q, t)$	$U_{qqtt} \leq 0$

In Table 1, Prospect 1 of the first row $(0.5, q-q_1; q-q_2)$ denotes a prospect where the subject has 50% probability to live with a QoL of $q-q_1$ for T years, and 50% to live in QoL of $q-q_2$ for T years. The other prospect of this first row is riskier, since it involves a lower minimum $(q-q_1-q_2)$ and a higher maximum (q). The other

prospects can be interpreted similarly. For cross-prudence and cross-temperance, $\tilde{t}$ and $\tilde{q}$, denote zero-mean risks on longevity and QoL, respectively.

In the model of Bleichrodt et al. (2003a), patients can choose the intensity *n* of a treatment combatting a disease. This only affects their QoL *q* and is risky, since it can either be effective, improving the patient's health by *b***n*, or it can be detrimental due to side effects, in which case the patient's health will deteriorate by *c***n*.[2] Hence, the amount of upside and downside potential depends on the treatment intensity chosen by the patient; the higher the intensity, the more extreme the outcomes will be. In this study we test the predictions of Bleichrodt et al. (2003a) by asking subjects to choose the amount *n* in this decision context, for different life durations *t*. For instance, in one of the questions the subject had to choose *n* such that they would live 20 more years with q $(0.5, 60\% - 0.1 \times n, 20y; 60\% + 0.4 \times n, 20y)$, with *n* measured in percentages, and *b*=0.4, *c*=-0.1; e.g. *n*=50% would correspond to $(0.5, 55\%, 20y;\ 80\%, 20y)$. Repeating this for several durations *t*, we could test the correlation with the risk traits from Table 1.

3. Experiment

3.1. Subjects

Participants were recruited randomly through a faculty internal recruitment system available to all undergraduate business students at the Rotterdam School of Management. As an incentive for taking part, participants were awarded with course credits. On arrival at the laboratory, a maximum of four students completed the procedure in the same room. A total of 124 students took part in the study. For two subjects, a program failure occurred during data collection. One student re-contacted us, asking to be excluded from the study because he had not answered faithfully. Therefore, a total of three cases were excluded from the study. The final sample size was N = 121 (51.2% female). The average age of participants was 20.1 years (SD = 1.44). n = 19 participants reported a physical health condition (16.0%), and n = 7 a mental health

[2] Note that this corresponds to therapeutic risk in the terminology of Eeckhoudt (2002) and Felder (2020).

condition (5.8%), and the average self-reported quality of life on the visual analogue scale ranging from 0 (death) to 100 (best possible health) was 83.48 (SD = 9.57). The average BMI was 21.52 (SD = 2.26), and n = 13 participants were considered underweight (10.7%), while n = 9 were considered overweight (7.4%).

3.2. Procedure

Ethical approval for this study was obtained from the ethical review board of Erasmus University. Subjects were first asked to provide their informed consent and signed a form of solemn commitment. Signing such a solemn commitment has been shown to increase diligent responding (Jacquemet et al., 2018, 2019). Subsequently, subjects received instructions to complete a part eliciting their risk attitudes and treatment proneness and completed 5 practice questions (1 for risk aversion with respect to QoL, 1 for correlation attitude, 1 for cross-prudence, 1 for cross-temperance, and 1 for treatment intensity). The order of the tasks was randomized. Within each trait, questions were not interspersed to avoid subjects having to switch between tasks continuously. Within each part, the questions were randomized. At the end of this part, 4 questions were repeated in order to test consistency (1 for question on correlation attitude, 1 one cross-prudence for longevity, 1 on risk aversion for longevity and 1 for treatment intensity). The experiment was programmed in Matlab. A researcher was in the room with the participants during all sessions.

3.3. Stimuli

For all tasks, we took a QoL level of q=60% of full health to be the base QoL. For longevity, this base was t=40 life years. As a result, risk aversion for QoL was elicited by fixing longevity at 40 years while varying the variance of QoL. Likewise, risk aversion for longevity was assessed by fixing QoL at 60% while varying the variance of longevity between the options. A similar procedure was used for the other traits. Table 2 shows the stimuli for all traits.

Table 2. Stimuli for the risk apportionment tasks.

Task [a]	Trait	Prospect A	Prospect B
1	Risk aversion for QoL	$[(60\%-10\%,40y);(60\%-40\%,40y)]$	$[(60\%,40y);(60\%-50\%,40y)]$
2		$[(60\%-10\%,40y);(60\%-20\%,40y)]$	$[(60\%,40y);(60\%-30\%,40y)]$
3		$[(60\%-20\%,40y);(60\%-20\%,40y)]$	$[(60\%,40y);(60\%-40\%,40y)]$
4	Risk aversion for longevity	$[(60\%,40y-10y);(60\%,40y-20y)]$	$[(60\%,40y-30y);(60\%,40y)]$
5		$[(60\%,40y-10y);(60\%,40y-10y)]$	$[(60\%,40y-20y);(60\%,40y)]$
6*		$[(60\%,40y-5y);(60\%,40y-10y)]$	$[(60\%,40y-15y);(60\%,40y)]$
7	Correlation attitude	$[(60\%-40\%,40y);(60\%,40y-10y)]$	$[(60\%-40\%,40y-10y);(60\%,40y)]$
8		$[(60\%-20\%,40y);(60\%,40y-20y)]$	$[(60\%-20\%,40y-20y);(60\%,40y)]$
9*		$[(60\%-20\%,40y);(60\%,40y-10y)]$	$[(60\%-20\%,40y-10y);(60\%,40y)]$
10	Cross-Prudence for longevity	$[(60\%,40y-20y);(60\%\pm20\%,40y)]$	$[(60\%\pm20\%,40y-20y);(60\%,40y)]$
11		$[(60\%,40y-10y);(60\%\pm40\%,40y)]$	$[(60\%\pm40\%,40y-10y);(60\%,40y)]$
12		$[(60\%,40y-10y);(60\%\pm20\%,40y)]$	$[(60\%\pm20\%,40y-10y);(60\%,40y)]$
13	Cross-Prudence for QoL	$[(60\%-20\%,40y);(60\%,40y\pm20y)]$	$[(60\%-20\%,40y\pm20y);(60\%,40y)]$
14*		$[(60\%-20\%,40y);(60\%,40y\pm10y)]$	$[(60\%-20\%,40y\pm10y);(60\%,40y)$
15		$[(60\%-40\%,40y);(60\%,40y\pm10y)]$	$[(60\%-40\%,40y\pm10y);(60\%,40y)$
16	Cross-Temperance	$[(60\%\pm20\%,40y);(60\%,40y\pm20y)]$	$[(60\%\pm20\%,40y\pm20y);(60\%,40y)]$
17		$[(60\%\pm40\%,40y);(60\%,40y\pm10y)]$	$[(60\%\pm40\%,40y\pm10y);(60\%,40y)]$
18		$[(60\%\pm20\%,40y);(60\%,40y\pm10y)]$	$[(60\%\pm20\%,40y\pm10y);(60\%,40y)]$

Note: [a] choice task was repeated once as a consistency check

3.4. Treatment intensity

Treatment proneness was operationalized as the preferred treatment intensity. Here, participants were presented with a singular 50-50 lottery, in which each outcome represented a quality of life index *q* for a given duration of life *t.* At baseline (intensity of 0%, i.e. no treatment taken), the two lotteries were identical. The life duration was always exogenous; that is, the subject could not influence the life duration. The life duration was equal for both lottery outcomes, and it was either certain or associated with uncertainty. The former case represents the situation in which the comorbidity caused a known reduction in life duration, whilst in the second case the comorbidity

caused a riskier life duration. The subject could, however, influence the expected quality of life by choosing a preferred treatment intensity *n*, represented as a percentage ranging from 0 to 100, which subjects could choose from in steps of 2%. The treatment is associated with either benefits *b* (associated with one lottery outcome) or costs *c* (associated with the other lottery outcome). The size of the benefits and costs depends on the treatment intensity *n*. The higher the treatment intensity, the higher the potential benefits as well as the potential costs. We picked a ratio of b/c=4, and used three questions with a fixed duration (20, 30 and 40 years), and one question with a random duration of either *t*=10 or *t*=30 years, equally likely. An overview of the stimuli is provided in Table 3.

Table 3. Stimuli for the treatment intensity task

	Task 1	Task 2	Task 3	Task 4*	Task 5
Prospect in case of intensity 0%	$(60\%, 20y \pm 10y)$	$(60\%, 20y)$	$(60\%, 30y)$	$(60\%, 40y)$	$(60\%, 40y \pm 10y)]$
Prospect in case of intensity 100%	$(50\% \; or \; 100\%, 20y \pm 10y)$	$(50\% \; or \; 100\%, 20y)$	$(50\% \; or \; 100\%, 30y)$	$(50\% \; or \; 100\% \, , 40y)$	$(50\% \; or \; 100\% \, , 40y \pm 10y)]$

*repeated at the end

3.5. Analysis

Data analysis was performed in R (R Core Team, 2016). We used the amount of choices (out of 3) that are compatible with a given risk trait as our measurement of the strength of multi- and univariate risk preferences. In our analysis, a subject is classified according to a risk trait if the majority of her choices is consistent with that particular trait. Thus, for example, an individual is classified as being risk averse (seeking) if most of her choices are compatible with risk aversion (seeking). For each of these traits, we investigated whether people show a given risk preference or behave at random based on a chi-square test. At the aggregate level, we report the average percentage of choices over tasks compatible with each trait. We use Fisher exact tests to compare the classifications obtained for each trait. To assess the relation between the higher-order

risk preferences and treatment intensity, we used repeated-measure ANOVAs and Friedman tests. We also used Wilcoxon and Student t-tests for complementary analysis.

Bleichrodt et al. (2003a) show under which conditions of the higher-order derivations, treatment intensity varies with duration. They show that an increase (decrease) in treatment intensity with duration is predicted by a decrease (increase) in risk aversion to health status with duration. In addition, when treatment intensity increases (decreases) with duration the sign of the following ratio is positive (negative):

$$r = \frac{U_{qqt}U_q - U_{qq}U_{qt}}{U_q^2}. \quad (2)$$

Bleichrodt et al. (2003a) show that r corresponds to the responsiveness of (normalized) correlation aversion to changes in health status. Because the denominator of the fraction in Eq. 2 is always positive, its sign depends on the sign of the numerator, which gives an unambiguous sign only if particular combinations of higher-order risk traits are satisfied. For example, if a participant is cross-prudent for longevity, risk averse for QoL and correlation seeking we know that the fraction is positive, whilst it is negative for a participant who is risk averse for QoL, cross-imprudent for longevity and correlation averse. Instead, in case of cross-prudence for longevity, risk aversion and correlation aversion, we cannot make a prediction for the sign of the fraction without knowing the degrees of the higher-order derivatives (the degrees of correlation aversion, cross-prudence and risk aversion for quality of life). We test if our data generate an unambiguous sign by computing Eq. 2 using the signs of the median traits, as well as computing the sign of Eq. 2 for each participant separately.

4. Results

4.1. Consistency checks

To assess whether participants were consistent in their answers, four items were included twice in the experiment, measuring risk aversion for duration, correlation aversion, cross-prudence for QoL, and treatment intensity.

For binary choices, subjects made the same choice in 75.38 percent of the repeated choices. This rate is consistent with the usually observed consistency rates in experiments (Stott, 2006, Attema et al, 2019). We also found some variability in consistency between the different tasks.[3] For the treatment intensity choices, subjects made the same choice in 41.32 percent of the repeated choices. This percentage can be considered relatively low. Allowing for an error margin of 5 percentage points, the consistency rate increases to 53.72 percent. For an error margin of 10 percentage points, it raises to 67.77 percent.

4.2. Risk preferences

Table 4 shows the results on risk preferences. The first two columns show the aggregate results: the mean proportion of the three choices compatible with each trait and the associated standard deviation. The last two columns show the individual results. The third column corresponds to the classification of individuals, based on their risk preferences, and the fourth shows the p-value of a one-sided binomial test for comparison between the percentage of individuals and 50 percent. At the aggregate level, we performed a series of chi-squared tests to check whether the observed distribution of preferences deviated from the distribution that would be observed if subjects choose randomly. All tests show that choices were not made at random.

Table 4. Risk preferences: aggregate results and individual classification.

	Aggregate results		**Individual classification**	
	mean	standard deviation	proportion	p-value
Risk aversion, quality of life	66.39	9.10	67.77	<0.01
Risk aversion, longevity	74.38	5.03	79.34	<0.01
Correlation aversion	10.19	2.66	4.13	<0.01
Cross-prudence for quality of life	36.64	3.73	32.23	<0.01
Cross-prudence for longevity	27.27	7.06	23.14	<0.01

[3] For the tasks measuring risk aversion the consistency rate was equal to 76.03 percent, for correlation aversion it was 88.43 percent. For the task measuring cross-prudence, the consistency rate was equal to 61.67 percent.

Cross-temperance	39.94	5.63	33.06	<0.01

We found risk aversion to be the predominant pattern for both longevity and quality of life, with a large majority of the choices compatible with risk aversion in both cases. Figure 1 illustrates this point and shows the distribution of the number of risk averse choices for quality of life and for longevity. Figure 1 also shows the expected number of risk averse choices if participants chose at random.

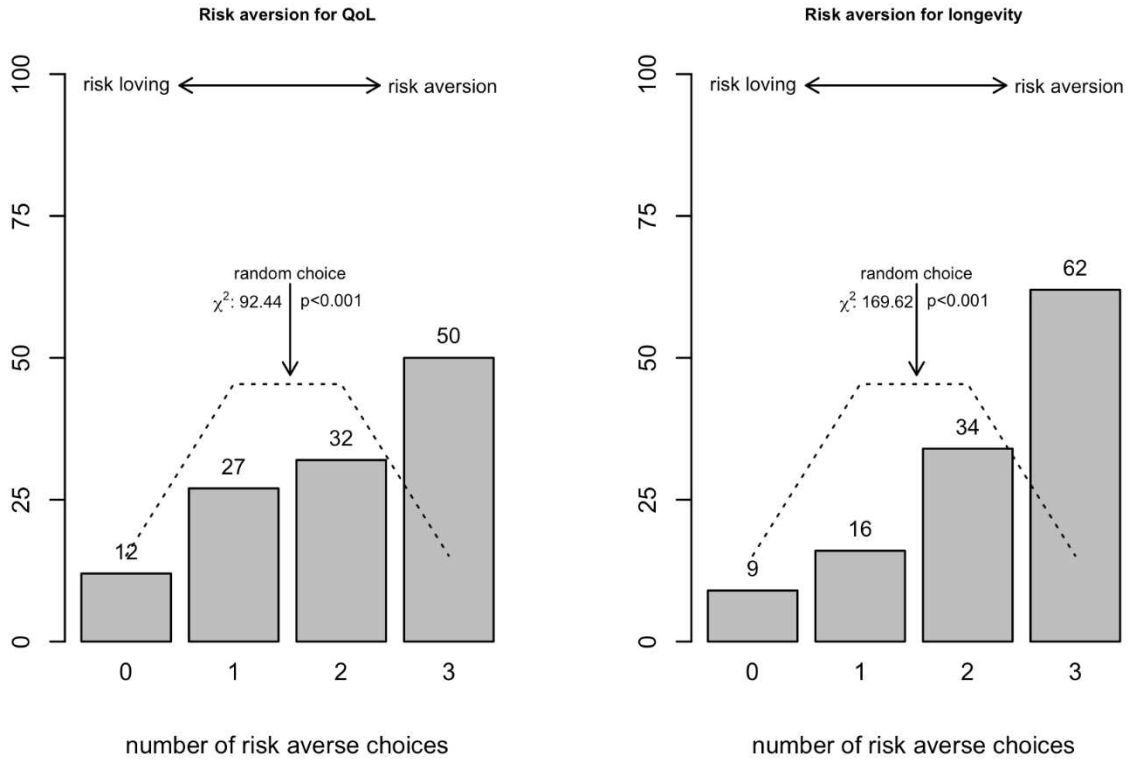

Figure 1: Distribution of the number of risk averse choices for quality of life and for longevity

Overall, 58.68 percent of individuals were classified as both risk averse for longevity and for quality of life. The association between risk attitudes for longevity and quality of life was highly significant (Fisher test, p-value 0.007).

We found a clear choice pattern indicative of a preference for correlation seeking for longevity and quality of life with more than 90 percent of the choices compatible with correlation seeking. Figure 2 shows the distribution of correlation averse choices for quality of life and for longevity together with the distribution of risk averse individuals

for quality of life (panel (a)) and for longevity (panel (b)). Under expected utility, this pattern of preference suggests that the cross-derivative of the utility function U_{qt} is positive for most individuals. Using classifications at the individual level, we found no evidence for an association between correlation attitudes and risk attitudes for neither quality of life nor longevity. Due to the large majority of individuals being classified as correlation seeking (95.87 percent), this result is hardly surprising.

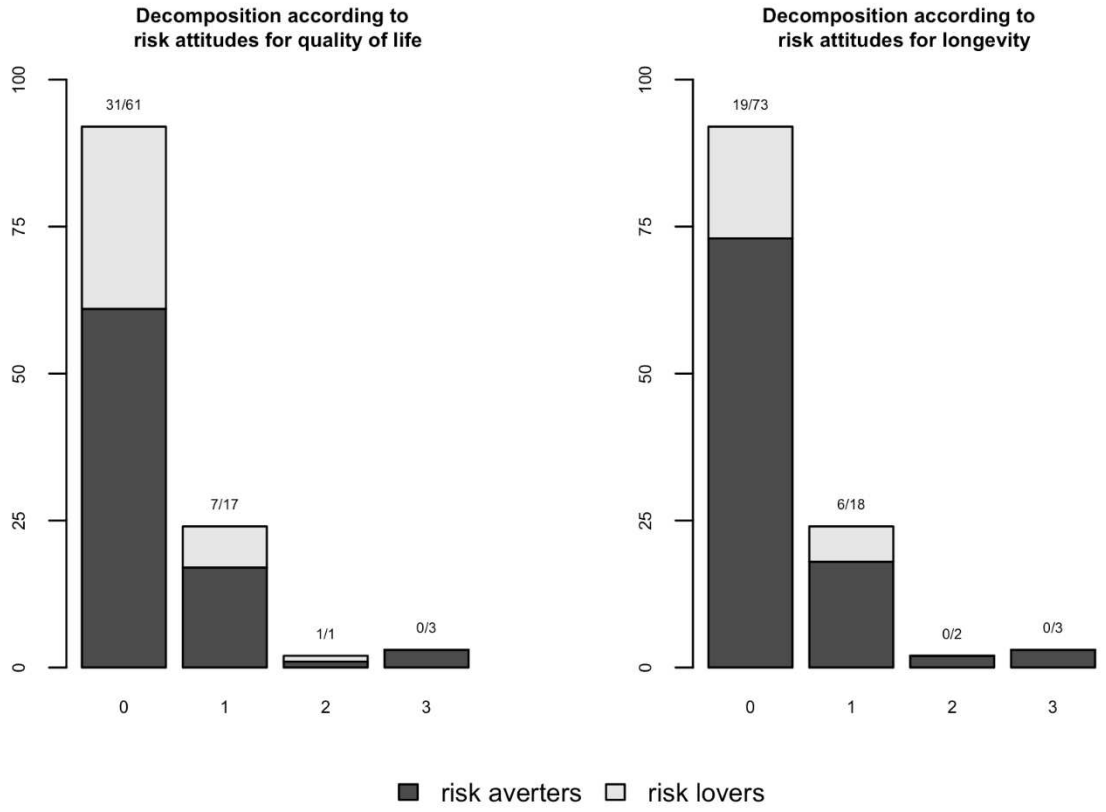

Figure 2: Correlation aversion and risk aversion for QoL and longevity, distribution of the number of correlation averse choices

Table 4 also shows the classification of individuals depending on their cross-prudent choices. A majority of individuals were classified as cross-imprudent for quality of life (67.77 percent of individuals) and for longevity (76.86 percent of individuals). At the individual level, 9.09 percent of individuals were classified as cross-prudent in both attributes and 53.72 percent as cross-imprudent for both attributes. We found no significant association between those risk preferences, and risk and correlation aversion. Under expected utility, the pattern of preferences revealed in Table 4 suggests

that the cross-derivatives of the utility function U_{qtt} and U_{qqt} were negative for most individuals.

Last, we found evidence for cross-intemperance with a majority of subjects choosing compatible with this trait. We found an association between cross-temperance and cross-prudence for quality of life (Fisher test, p-value 0.014) but not for longevity (Fisher test, p-value 0.82). The combination of correlation seeking, cross-imprudence and cross-intemperance corresponded to the modal multivariate risk preference when both cross-prudence for quality of life and longevity were considered (for 48.76 percent and 52.07 percent of subjects, respectively).

4.3. Choice of treatment intensity with certain longevity

Table 5 shows the descriptive statistics on the choice of treatment intensity. On average, subjects chose a treatment intensity of 60%. The values for the third quartile show that a significant number of individuals chose the maximum treatment intensity in any treatment.

Table 5. Descriptive statistics on the choice of treatment intensity.

	Certain longevity			**Risky longevity**	
	T=20	T=30	T=40	T=20+10/-10	T=40+10/-10
median	74.00	61.00	60.00	60.00	58.00
Q1	36.00	39.00	32.00	30.00	38.00
Q3	99.50	99.50	99.50	99.50	99.50
mean	64.41	63.12	62.14	58.32	59.74
sd	35.42	32.07	34.46	36.13	33.64

The median values reported in the first three columns of Table 5 suggest that treatment intensity decreases with longevity, while the means suggest a flat pattern instead. In order to test the association between treatment intensity and longevity, we ran a repeated-measure ANOVA with longevity as the within-subject factor. In

accordance with the mean values from Table 5, the results show that treatment intensity does not differ between the three tasks (p-value 0.72). A Friedman test shows however a marginally significant difference between the median values (p-value 0.08). Pairwise comparisons based on Wilcoxon or Student t-test support the results from the ANOVA.

Table 6 shows a classification of individuals based on the relation between longevity and treatment intensity. We used two rules to classify subjects. The strict rule classifies individuals as having a constant (increasing, decreasing) profile if they reported the same exact (increasing, decreasing) treatment intensity for the three longevities $T = 20,30,40$. We also used a more lenient rule allowing for a deviation of 5 percentage points in first-order differences. Subjects who were classified as neither constant nor increasing or decreasing were classified as exhibiting a non-monotone profile. Individual analysis from Table 6 shows that between $1/4$ and $1/3$ of the subjects chose constant treatment intensities for different longevities, a majority of them choosing extreme (0 and 100 percent) treatment intensities. For around $1/4$ of the subjects, treatment intensity decreases with longevity and for around $1/6$ of the subjects, treatment intensity increases with longevity.

Table 6. Classification of individuals depending on the relationship between treatment intensity and longevity.

	Strict rule	With 5 pp. error
Constant	32	40
Constant with extreme choices	27	27
Decreasing	32	33
Increasing	21	23
Non-monotone	36	25

We now use the classification of individuals based on their choice apportionment in binary choice to evaluate the prediction of the expected utility model. More specifically,

we use the individual classification of risk aversion for quality of life, correlation aversion and cross-prudence for longevity to infer the sign of r defined in Eq (2). Remember that r measures the responsiveness of (normalized) correlation aversion to change in health status and is the key behavioral parameter governing the response of treatment intensity to duration.

For 60.33 percent of the individuals, the information gathered from binary choices did not allow to have a clear prediction on the sign of r. 25.62 percent were classified as revealing a negative r. The remaining 14.05 percent were classified as revealing a positive r. Because the test of expected utility is based on the sign of r, the risk apportionment technique does not allow to make firm predictions for a majority of subjects in our experiment. Figure 3 shows the distribution of treatment intensities at different longevities, based on the revealed sign of r. A visual inspection of Figure 3 shows that treatment intensities tend to decrease for participants revealing a positive r and a non-monotone pattern for those revealing a negative r. An ANOVA with repeated-measures for subjects with a revealed negative r cannot reject constancy of treatment intensity (p-value 0.45). The same applies for the mixed case but also for positive r. In accordance with Figure 3, for the two latter classifications, a Friedman test nevertheless shows marginally significant differences (p-value of the Friedman test 0.08 and 0.05).

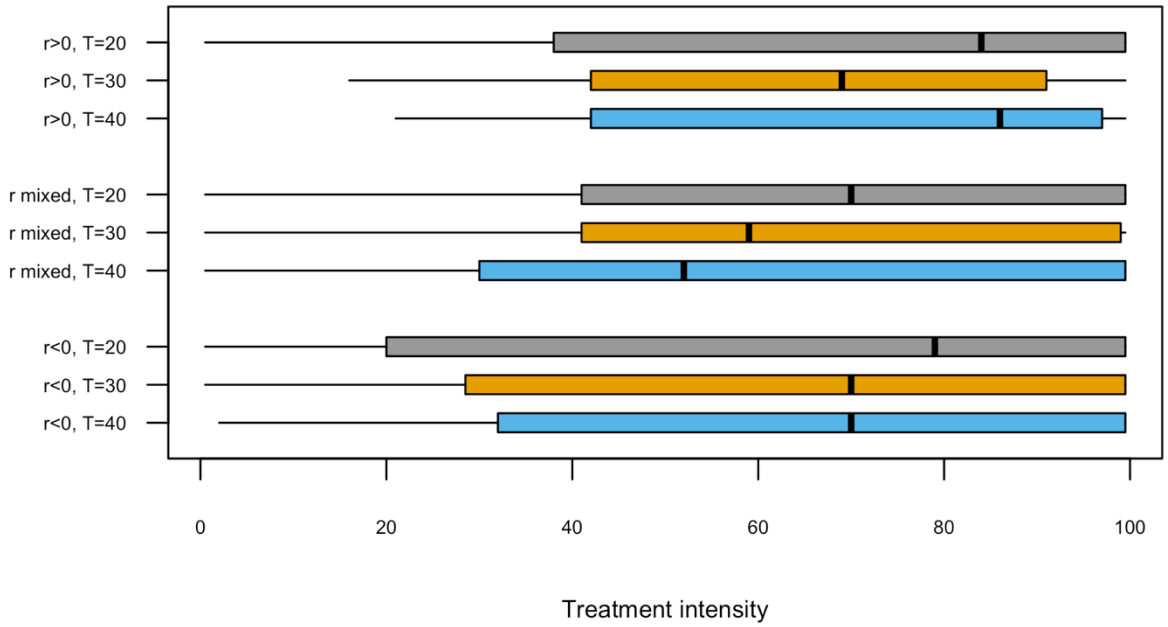

Figure 3: Relations between treatment intensity and the sign of responsiveness of normalized correlation attitude to changes in health status r.

4.4. Choice of treatment intensity with risky longevity

The values reported in Table 5 suggest that treatment intensity decreases when a risk on longevity is introduced. In order to test for the impact of risky longevity on treatment intensity, we run a repeated-measure ANOVA, with two within-subject factors (certain vs. risky longevity and expected longevity equal to either $T = 20$ or $T = 40$). The results from the ANOVA show that treatment intensity did not vary with longevity (p-value 0.89) and that riskiness of longevity has only a marginal impact on treatment intensity (p-value 0.06). Pairwise comparisons based on Wilcoxon or Student t-tests support the results from the ANOVA: the differences between treatment intensities at certain and risky longevity were not significantly different at expected longevity equal to $T = 40$ years, but were marginally different at expected longevity equal to $T = 20$ years (Wilcoxon two-sided test, p-value 0.07, Student two-sided t-test, p-value 0.07). One-sided interpretations of pairwise comparisons therefore show evidence for decreasing

treatment intensity with riskiness of longevity, at least when expected longevity is low. Last, the base value of treatment intensity (t=20 or t=40) did not impact treatment intensity when duration was risky[4] (Wilcoxon two-sided test, p-value 0.49, Student two-sided t-test, p-value 0.61).

Bleichrodt et al. (2003a) show that risk aversion alone is not sufficient to predict the reaction of the introduction of a risky longevity in the choice of treatment intensity. In particular, they show that it is far from obvious that the riskiness of longevity leads, through risk aversion, to a decrease in treatment intensity. We tested this hypothesis by comparing the differences between risky and certain longevity (for expected longevity equal to either t=20 or t=40) for risk averse and risk seeking subjects. Results are shown in Figure 4, which makes it clear that risk aversion, per se, is not clearly associated with a systematic drop in treatment intensity when longevity is risky[5]. Figure 4 also shows that risk seeking does not translate in a systematic increase in treatment intensity when risky longevity is introduced. A repeated-measure ANOVA, with one within-subject factor (expected longevity equal to either t=20 or t=40) and one between-subject factor (risk attitudes), shows that the former has no significant effect on the difference between treatment intensity for risky or certain longevity (p=0.156). Together, these results confirm Bleichrodt et al. (2003a) that there is not a one-to-one link between risk aversion for longevity and choice of treatment intensity when longevity becomes risky.

[4] At the individual level, we used a classification similar to the one shown in Table 3. We classified participants based on the relation between riskiness of longevity and treatment intensity using a strict rule and a more lenient rule with 5 percentage points tolerance on first-order differences. Results show that between 1/4 and 1/3 of the subjects chose constant treatment intensity at different longevities. For around 1/4 of the subjects, the treatment intensity increases with risky longevity and for around 1/6 of the subjects, treatment intensity decreases with risky longevity.

[5] In Figure 4, the variation in treatment intensity corresponds to the difference in treatment intensity between risky and certain longevity. A negative value indicates that the chosen treatment intensity was higher under certain than under risky longevity.

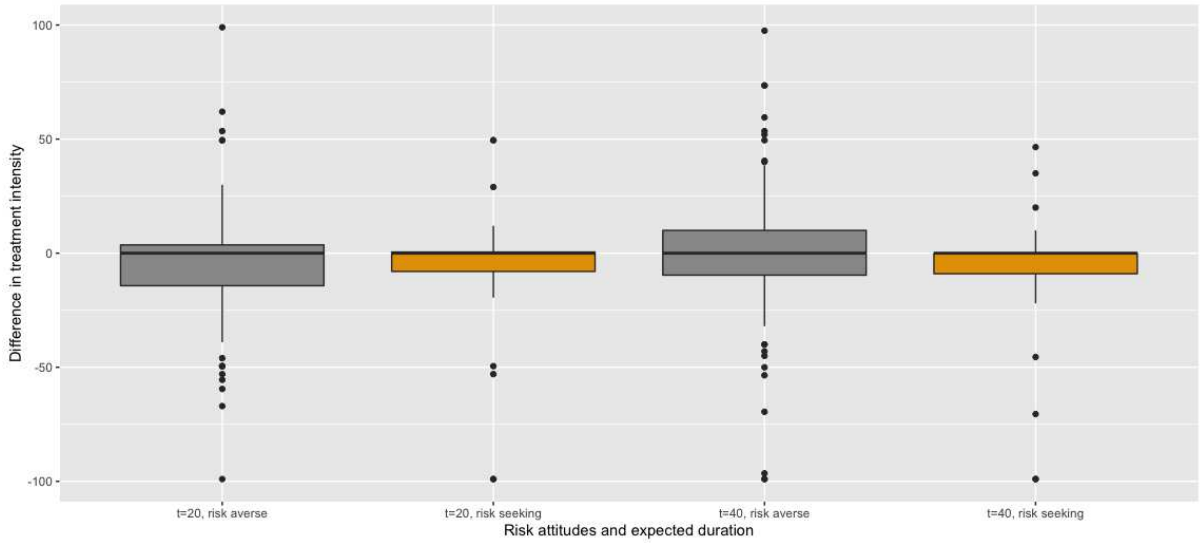

Figure 4: Relations between variation in treatment intensity and risk attitudes

5. Discussion

The study set out with two objectives. First, we aimed to describe people's multivariate and higher-order risk attitudes for longevity and quality of life. Second, we conducted a test of the QALY model by assessing how people's higher-order risk attitudes were related to their preference for treatment intensity.

Our findings for the risk apportionment task confirm the intuitive predictions of Bleichrodt et al. (2003a) that people are risk averse and correlation seeking for duration and quality of life. Concerning risk aversion, this is a reassuring finding, in accordance with previous evidence (Attema et al., 2013, 2016, Delprat et al., 2016). The finding of correlation seeking on the other hand is particularly interesting given the widespread evidence of correlation aversion for other outcomes (Ebert and van de Kuilen, 2015), although it has been found for the QALY model before (McNeil et al., 1981; Pliskin et al., 1980; Sutherland et al., 1982). Under expected utility, correlation seeking reveals that increasing longevity reinforces the marginal utility of variations (positive or negative) in quality of life insofar as individuals benefit from it or experience it longer. This study's results also indicate a clear majority of cross-imprudent choices, albeit less deviant from neutrality than for the second-order traits. Lastly, the evidence is, as usual, least pronounced for intemperance, but still we found a significant deviation from 50%. The model of Bleichrodt et al. (2003a) neither provides

any predictions for the signs of these higher-order preferences, nor does it give intuitive predictions. Hence, our study provides the first evidence of these higher-order, multivariate, risk preferences. These findings can have large implications for several health-related behaviors and open up a new research area.

We found no correlation between longevity and treatment intensity (i.e. health status risks). This result is in contrast with the prediction of Bleichrodt et al. (2003a), who argued that it would be plausible for health status risks to decrease with life expectancy. The absence of such a relation suggests that (this part of) the QALY model is valid, because it implies comorbidities that only affect life duration indeed have no impact on treatment decisions that only affect quality of life. However, we admit that there are several caveats to this conclusion. First, the treatment intensity task may not have been the best way to elicit treatment preferences, which could explain the high amount of noise and the multimodal preferences observed in this task. Second, since the theoretical analysis by Bleichrodt et al. (2003a) assumes expected utility, our test of the QALY model based on their framework is only valid to the extent that expected utility holds. Otherwise, it may be the case that the observed findings are due to a falsification of expected utility, and that the QALY model would not be valid in a non-EU framework. It is left to future research to test these properties of the QALY model without the restriction to EU. Still, our results regarding the sign of Eq. 2 do not contradict the lack of a correlation between treatment intensity and duration and duration risk, lending some credibility to the test.

We report a differential impact of the introduction of a background risk on longevity for different amounts of longevity. For high expected longevity, the background risk did not impact the choice of treatment intensity while it significantly decreased it for lower expected longevity. According to this result, the QALY model, which imposes a neutral impact of the background risk on treatment decisions, appears to be more robust for long durations than for short durations. This result confirms the empirical results from Bleichrodt et al. (2003b) and Attema and Brouwer (2008), who showed in a very different experimental setting that standard elicitations of the QALY model (standard gamble or time trade-off) are more likely to be biased for short durations than for long durations.

The use of risk apportionment techniques to identify higher-order risk attitudes, and therefore infer the properties of the utility function has its own limitations. Under expected utility, risk apportionment techniques allow to obtain clear measurements of the signs of successive derivatives of the utility function from behavioral traits. The method is easy to handle for experimenters and the elements of choices are rather easy to understand for participants to an experiment. However, risk apportionment techniques perform poorly if one needs to obtain precise knowledge on the shape of the utility function. Such knowledge is required if one wants a precise elicitation of the effect of a risky comorbidity on the optimal treatment decision. For such comparative statics results, elicitation of risk aversion and other higher-order risk preferences by risk apportionment techniques are too coarse to elicit all the determinants of marginal benefits and marginal costs of treatment. A precise empirical assessment of those comparative statics would require an elicitation of more complex objects such as prudence premia for longevity and health status.

Another limitation is that we used a student sample for our lab study. Although this sample is not representative of the general public, it was useful for a first test application of risk apportionment techniques to the QALY model. Nevertheless, a clear drawback of our young sample is that they are unlikely to have much experience with illness. Hence, our conclusions, even if firm (especially for correlation seeking), should be interpreted with caution and future research should test if our first results can be generalized to the general public's preferences and, perhaps, patient preferences.

Third, given that we studied preferences for health, we could obviously not use real incentives. However, we are not aware of any way to implement real incentives for this type of preferences, so the use of hypothetical incentives seems to be the best currently available alternative. A large literature has been developed on willingness to pay for health gains and QALY gains (Ryen and Svensson, 2015) and could be used as a basis for attaching financial incentives to health gains and losses. The methods developed in that literature might not be suitable to implement financial incentives in an experimental setting for several reasons. First, the monetary amounts attached to QALY gains (and losses) are rather substantial and incompatible with experimental implementation. As an example, Bobinac et al. (2014) report willingness to pay for one QALY gains to range from 80,000 euros to 250,000 euros, depending on the underlying theoretical

assumptions. Second, most studies estimate willingness to pay under certainty, which creates an upward bias in the estimation (Olofsson et al., 2019). Last, most of the literature on willingness to pay for health gains and QALY gains is based on hypothetical tasks and incentives are mostly devoted to reward participation.

The QALY model has been largely challenged as a descriptive model for health decisions, mainly because of violations of expected utility (Bleichrodt and Pinto, 2005). One of the reasons why the QALY would fail to represent risk preferences is therefore largely due to biases and heuristics in elicitation methods, such as the certainty effect (Bleichrodt et al., 2007) or loss aversion (Bleichrodt et al., 2003b). In this paper we used a different methodology, based on risk apportionments, to assess the descriptive ability of the QALY model. One advantage of this methodology rests on its use of paired gambles, for which expected utility is less likely to be violated. Our results show that, at least within expected utility, the QALY model could not be easily rejected.

References

Attema, A. E., & Brouwer, W. B. (2008). Can we fix it? Yes we can! But what? A new test of procedural invariance in TTO-measurement. *Health Economics*, *17*(7), 877-885.

Attema, A. E., Brouwer, W. B., & L'Haridon, O. (2013). Prospect theory in the health domain: A quantitative assessment. *Journal of Health Economics*, *6*(32), 1057-1065.

Attema, A. E., Brouwer, W. B., L'Haridon, O., & Pinto, J. L. (2016). An elicitation of utility for quality of life under prospect theory. *Journal of Health Economics*, *48*, 121-134.

Attema, A.E., L'Haridon, O., & van de Kuilen, G. (2019). Measuring Multivariate Risk Preferences in the Health Domain. *Journal of Health Economics*, *64*, 15–24.

Bleichrodt, H., Abellan-Perpiñan, J. M., Pinto-Prades, J. L., & Mendez-Martinez, I. (2007). Resolving inconsistencies in utility measurement under risk: Tests of generalizations of expected utility. *Management Science*, *53*(3), 469-482.

Bleichrodt, H., Crainich, D., & Eeckhoudt, L. (2003a). The effect of comorbidities on treatment decisions. *Journal of Health Economics*, *22*, 805–820.

Bleichrodt, H., & Pinto, J. L. (2005). The validity of QALYs under non-expected utility. *The Economic Journal*, *115*(503), 533-550.

Bleichrodt, H., Pinto, J. L., & Abellan-Perpiñan, J. M. (2003b). A consistency test of the time trade-off. *Journal of Health Economics*, *22*, 1037-1052.

Bobinac, A., van Exel, J., Rutten, F. F., & Brouwer, W. B. (2014). The value of a QALY: individual willingness to pay for health gains under risk. *Pharmacoeconomics*, *32*(1), 75-86.

Courbage, C., & Rey, B. (2006). Prudence and optimal prevention for health risks. *Health Economics, 15*, 1323–1327.

Delprat, G., Leroux, M. L., & Michaud, P. C. (2016). Evidence on individual preferences for longevity risk. *Journal of Pension Economics & Finance*, *15*(2), 160-179.

Ebert, S., & van de Kuilen, G. (2015). Measuring multivariate risk preferences. *Available at SSRN 2637964*.

Eeckhoudt, L. (2002). *Risk and medical decision making, Vol. 14*. Springer Science & Business Media.

Eeckhoudt, L., Rey, B., & Schlesinger, H. (2007). A good sign for multivariate risk taking. *Management Science, 53*, 117–124.

Eeckhoudt, L., & Schlesinger, H. (2006). Putting Risk in Its Proper Place. *American Economic Review, 96,* 280–289.

Felder, S. (2020). The treatment decision under uncertainty: The effects of health, wealth and the probability of death. *Journal of Health Economics, 69*, 102253.

Jacquemet, N., Luchini, S., Rosaz, J., & Shogren, J.F. (2018). Truth telling under oath. *Management Science, 65*, 426–438.

Jacquemet, N., Luchini, S., Shogren, J.F., & Watson, V. (2019). Discrete Choice under Oaths. *Working paper*.

Keeney, R.L., & Raiffa, H. (1993). *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge university press.

Krieger, M., & Mayrhofer, T. (2012). Patient Preferences and Treatment Thresholds under Diagnostic Risk – An Economic Laboratory Experiment. *Rheinisch-Westfälisches Institut für Wirtschaftsforschung, Ruhr-Universität Bochum, Universität Dortmund, Universität Duisburg-Essen.*

MacMahon, S. (2018). *Multimorbidity: A priority for global health research*. Academy for Medical Science, London, UK.

McNeil, B.J., Weichselbaum, R., & Pauker, S.G. (1981). Speech and Survival: tradeoffs between quality and quantity of life in laryngeal cancer. *New England Journal of Medicine, 305*, 982–987.

Olofsson, S., Gerdtham, U. G., Hultkrantz, L., & Persson, U. (2019). Value of a QALY and VSI estimated with the chained approach. *The European Journal of Health Economics*, *20*(7), 1063-1077.Pauker, S.G. (2014). Moments When Utilities Are Functional. *Medical Decision Making, 34*, 4–7.

Pliskin, J.S., Shepard, D., & Weinstein, M.C. (1980). Utility functions for life years and health status. *Operations Research, 28,* 206–224.

Ryen, L., & Svensson, M. (2015). The willingness to pay for a quality adjusted life year: a review of the empirical literature. *Health Economics*, *24*(10), 1289-1301.

Sanders, G.D., Neumann, P.J., Basu, A., Brock, D.W., Feeny, D., Krahn, M., Kuntz, K.M., Meltzer, D.O., Owens, D.K., Prosser, L.A., Salomon, J.A., Sculpher, M.J., Trikalinos, T.A., Russell, L.B., Siegel, & J.E., Ganiats, T.G. (2016). Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. *Journal of the American Medical Assocation, 316*, 1093–1103.

Sutherland, H.J., Llewellyn-Thomas, H., Boyd, N.F., & Till, J.E. (1982). Attitudes Toward Quality of Survival. *Medical Decision Making, 2*, 299–309.

Appendix

Table 7 shows the percentages of choices compatible with risk apportionment for each binary choice task in the experiment.

Table 7: Number and percentage of risk apportionment choices for each task

	Task	**#**	**percentage**	**% consistency**
Risk aversion for quality of life	1	73	60.33	
	2	75	61.98	
	3	93	76.86	
Risk aversion for longevity	1	87	71.90	76.03
	2	86	71.07	
	3	97	80.17	
Correlation aversion	1	10	8.26	88.43
	2	11	9.09	
	3	16	13.22	
Cross-prudence for quality of life	1	44	36.36	61.67
	2	40	33.06	
	3	49	40.50	
Cross-prudence for duration	1	34	28.10	
	2	24	19.83	
	3	41	33.88	
Cross-temperance	1	56	46.28	
	2	43	35.54	
	3	46	38.02	