Risk attitudes in medical decisions for others: an experimental approach

Arrieta, Alejandro; Garcia Prado, Ariadna; Gonzalez, Paula; Pinto Prades, Jose Luis

Published in:
Health Economics

DOI:
10.1002/hec.3628

Publication date:
2017

Document Version
Peer reviewed version

Health Economics

Risk Attitudes in Medical Decisions for Others: An Experimental Approach

Journal:	*Health Economics*
Manuscript ID	HEC-17-0105.R3
Wiley - Manuscript type:	Supplement paper
Keywords:	medical decisions for others, risk aversion, health contexts, laboratory experiment, multiple price list format

Risk Attitudes in Medical Decisions for Others: An Experimental Approach*

ABSTRACT

The aim of this paper is to investigate how risk attitudes in medical decisions for others vary across health contexts. A lab experiment was designed to elicit the risk attitudes of 257 medical and non-medical students by assigning them the role of a physician who must decide between treatments for patients. An interval regression model was used to estimate individual coefficients of relative risk aversion, and an estimation model was used to test for the effect of type of medical decision and experimental design characteristics on elicited risk aversion. We find that: (i) risk attitudes vary across different health contexts, but risk aversion prevails in all of them; (ii) students enrolled in health-related degrees show a higher degree of risk aversion; and (iii) real rewards for third parties (patients) make subjects less risk-averse. The results underline the importance of accounting for attitudes towards risk in medical decision-making.

Keywords: medical decisions for others, risk aversion, health contexts, laboratory experiment, multiple price list format.

JEL Codes: I1, C91, D81

*We appreciate valuable comments by Florian Artinger, Irene Gayarre, Glenn Harrison, Heike Hennig-Schmidt, Graham Loomes, Jon Benito Ostolaza, Emil Persson, Daniel Wiesen and participants at the 4th Workshop in Behavioral and Experimental Health Economics (Cologne). We also acknowledge Nathaniel Barrett's support in refining the text. Financial support from Ministerio de Ciencia y Tecnología and FEDER (ECO2012-3648, ECO2013-43526-R, ECO2015-65408-R and ECO2015-65031-R) and Junta de Andalucía (SEJ-04992) is gratefully acknowledged. We do not have conflict of interests nor unethical background to declare.

1. **Introduction**

Risk preference has been shown to influence a variety of behaviors under conditions of uncertainty. As would be expected, given the prevalence of such conditions surrounding questions of healthcare provision, there is an extensive literature on risk attitudes in the health sector, most of which focuses on measuring patient preferences and propensity towards risk. Physicians' risk attitude is also an important element of service provision, but to date it has received comparatively little attention. The question of whether to administer one treatment or another to a patient is a problem confronted by physicians on a daily basis. In such situations, the physician's attitude towards risk may affect diagnosis patterns and treatment recommendations (Nightingale and Carter, 1987; Sommers and Zeckhauser, 2008; Chandra et al., 2012; Michel-Lepage et al., 2013; Massin et al., 2015 and Schosser et al., 2016).[1] Thus, a study that elicits risk preferences in medical decisions for others could contribute to a better understanding of the diagnostic and therapeutic decisions that direct medical practice.

In addition, the stability of risk preferences is a crucial question among economists, psychologists and behavioral researchers. Economists have generally assumed that individuals exhibit a single risk preference that governs risk-taking behavior in all circumstances. However, this has been debated within the psychology literature, as there is evidence that risk attitudes vary across different outcomes and domains (MacCrimmon and Wehrung, 1990; Weber et al., 2002, Hanoch et al., 2006). In fact, recent economic papers have also acknowledged the existence of domain-specific risk preferences (Barseghyan et al., 2011, Einav et al., 2012, Galizzi et al, 2016a). However, with the exception of van der Pol and Ruggeri (2008), Ruggeri and van der Pol (2012) (hereafter VdP&R, 2008, 2012) and Schosser et al. (2016), we are not aware of studies that have analyzed whether risk attitudes vary across different contexts within the health domain.[2]

[1] For instance, Michel-Lepage et al. (2013) and Massin et al. (2015) show how risk aversion affects physicians' decisions regarding vaccination and the use of antibiotics for treating streptococcal infections, respectively.

[2] Butler et al. (2012) also acknowledge such differences and develop a medical risk domain subscale for DOSPERT—the most frequently used test to differentiate risk attitudes across different domains—that includes blood donation, kidney donation, daily medication use for allergies, knee replacement surgery, general anesthesia in dentistry, and clinical trial participation as items.

With these questions in mind, this paper contributes to this scarce literature on risk attitudes in the health domain by applying an experimental economics approach. The use of controlled laboratory experiments in health economics is growing (Schram and Sonnemans, 2011; Buckley et al., 2012; Godager and Wiesen, 2013; Brosig-Koch et al.,2016; Galizzi and Wiesen, in press), as these experiments enable researchers to investigate health topics for which field evidence is difficult to obtain.

We carried out a lab experiment with 257 students by assigning them the role of a physician who must decide between treatments for patients in different contexts. Our aim was to investigate whether the risk attitudes of these subjects vary across outcomes within the health domain. Our investigation focused on risk attitudes when deciding for others, i.e., the students acted as doctors prescribing treatments for their patients. Risk attitudes in "decisions for others" have received much less attention in the health domain than in decisions for oneself. However, there is evidence that physicians recommend different treatments for patients than they would choose for themselves (Ubel et al., 2011; García-Retamero and Galesic, 2012).[3]

In order to investigate whether risk attitudes are context-dependent, we introduced three health contexts that correspond to three different types of medical decisions: a context with outcomes expressed as years of life in perfect health, a context of a terminal illness with outcomes expressed as months of life in moderate-ill health, and a context with outcomes expressed as hours of pain alleviated after surgery.

We run our experiment with medical students as well as students enrolled in other non-medical degrees, which allow us to approximate physicians' decisions and to test for the effects of university degree (health related versus non-health related) on treatment choices.

Finally, it is important to note that most of the health literature on preference elicitation has used hypothetical questions (see Galizzi et al., 2016a for a review). In the money domain there is mixed evidence about whether risk attitudes differ depending on whether hypothetical or

[3] Similar patterns of behavior have been documented in studies in the money domain, although the results of these studies on risk aversion are mixed. Sutter (2009) and Chakravarty et al (2011) find increased risk taking on behalf of others, while Reynolds et al. (2009), Eriksen and Kvaloy (2010) and Anderson et al. (2016) find the opposite result.

real rewards are used. While some find that risk attitudes differ between hypothetical and money-rewarded questions (Holt and Laury 2002, 2005 -hereafter H&L, 2002, 2005-; Harrison and Rutström, 2008b), others suggest that there is no such difference (Camerer and Hogarth, 1999; Hertwig and Ortmann, 2001; Bardsley et al., 2010; Abdellaoui et al., 2013; Cohen et al., 2016). As this question is still being debated, and evidence from health experiments with real rewards is scarce, we are interested to determine whether elicited preferences for health treatments differ between purely hypothetical frameworks and frameworks where real rewards are at stake. For this purpose, participants in the real group were informed that the money obtained as a result of their medical decisions would be transferred to an internationally known non-profit institution that provides medical services for the poorest populations in India.

In order to elicit preferences, we used a multiple price list (MPL) method, using the well-known test developed by H&L in 2002. The H&L method has been widely used in experimental economics to evaluate risk preferences in a variety of contexts (Andersen et al., 2008; Charness and Viceisza, 2017; Dohmen and Falk, 2011; see Philippin and Crosetto, 2016 for a review on H&L experiments).[4] However, it has been scarcely applied in the health domain, where standard gamble methods are the most commonly used methods to elicit preferences towards health risks. The exceptions are Galizzi et al. (2016a) and Galizzi et al. (2016b), which adapt the questions of H&L (2002) to the health domain in order to analyze whether patients' risk preferences differ across the health and financial domains and whether risk preferences differ between patients and their doctors. The H&L method presents a structured menu of binary choices, where lotteries are compared in pairs, thereby providing interval estimates of risk aversion. The advantage of H&L method is that it has risk in both lotteries (Farquhar, 1984), which avoids the "certainty effect"[5] identified as a weakness of the Standard Gamble and makes it easier for subjects to understand and choose an option.[6] Moreover, there is evidence

4 The prevalent use of the Holt–Laury measure has allowed researchers to compare risk attitudes across a wide array of contexts and environments. In turn, this has facilitated a less fragmented approach to the study of risk preferences that minimizes methodological differences and aims to characterize a more general phenomenon (Charness et al. 2013).

5 Standard gamble compares a lottery with a certain option. This can overweight the certain option as outlined by Kahneman and Tversky (1979).

6 In addition, as Holt and Laury (2013) state: "having lotteries for both options controls for a difference in complexity, i.e. the time required to determining the payoff, or a preference for playing or avoiding the lottery independently of the probabilities involved."

suggesting that Expected Utility is more descriptively valid when subjects compare pairs of non-degenerate gambles (Bleichrodt et al., 2007; Abdellaoui et al., 2011). Since the H&L method can potentially result in a bias toward the choice of options located in the middle of the table where risk neutrality is defined (Andersen et al., 2006; Harrison et al., 2007a; Harrison and Rutström, 2008a; Lévy-Garboua et al., 2012; Charness et al., 2013), we followed Andersen et al. (2006) and developed two asymmetric frames that are skewed in opposite directions.

Our results confirm that risk preferences are context dependent. Although risk aversion dominates in all types of medical decisions, participants in the experiment seem to exhibit more risk aversion in the context of hours of pain alleviated. The finding that risk aversion differs according to context corroborates previous studies (VdP&R, 2008, 2012) but the finding that our subjects are clearly risk averse in all contexts does not. Our findings also show that students who are enrolled in health-related degrees were more risk averse than students from other degrees. Finally, we find that introducing real rewards that affect others (i.e. patients) seems to make subjects less risk averse. Regarding the asymmetric frames, biases or "skewness effects" are small and do not affect the main conclusions of the paper.

The paper is organized as follows. In the following section we present our research questions. Section 3 describes the experimental design and procedures for collecting data. Section 4 explains the methods followed. Section 5 presents the results. Finally, Section 6 discusses the results and concludes with suggestions for further research.

2. Research Questions

In addition to eliciting preferences toward risk in medical decisions for others, this paper seeks to address three research questions:

Research Question 1: Is risk attitude context-dependent in the health domain, i.e., is risk aversion constant for all types of medical decisions?

As mentioned in the previous section, recent economic papers have acknowledged the existence of domain-specific risk preferences. For instance, households are more risk averse when making choices about home insurance as compared to car insurance (Barseghyan et al., 2011). Or, similarly, individuals are less risk averse in the financial domain as compared to the health domain (Galizzi et al, 2016a). Based on these findings, it seems likely that different risk attitudes can be found across different kinds of medical problems (see also Nightingale and Carter 1987; Michel-Lepage et al. 2013, Massin et al 2015 and Schosser et al 2016). Using the certainty equivalent method and the probability equivalent method, VdP&R (2008, 2012) find that risk attitudes in the health sector are indeed context dependent: individuals tend to be risk averse with respect to gambles that involve risk of immediate death and risk seeking in gambles involving quality of life, although they find conflicting evidence in the case of gambles involving life expectancy without risk of immediate death.[7] In this paper we aim to provide further evidence concerning the variation of risk attitude across medical contexts, in this case using a different elicitation method procedure based on H&L's measure of risk aversion. Accordingly, in the experiment we designed three different health contexts that had not been previously analyzed for this purpose in the literature.

Research question 2: Are students enrolled in health-related degrees different from others regarding risk aversion?

The experimental literature has shown that there are differences of risk preference across different subgroups, as found between men and women (Byrnes et al., 1999), between young and old (Harrison et al., 2007b), and between novices and experts (Dyer et al. 1989). Moreover, different papers show that medical students behave differently in areas such as other-regarding motivation and moral behavior (Hennig-Schmidt and Wiesen, 2014; Jagsi and Lehmann, 2004). Thus, it seems reasonable to expect that risk preferences differ between medical students (future physicians) and non-medical students since the former have higher awareness of the benefits and risks involved in health treatments. In our experiment students from health and non-health degrees were recruited in a proportion of 2 to 1.

[7] In the 2008 paper they found that subjects were risk seeking in this case while in the 2012 paper they were risk averse.

Research question 3: Do real rewards play a role in eliciting preferences for treatments if the payments affect a third party?

Several well-known studies in the literature that compare hypothetical versus money-rewarded questions have yielded results indicating a higher degree of risk aversion for the group with real rewards (H&L, 2002, 2005; Harrison and Rutström, 2008b). Others, however, do not find significant differences (Camerer and Hogarth, 1999; Hertwig and Ortmann, 2001; Bardsley et al., 2010; Abdellaoui et al., 2013; Cohen et al., 2016). However, all of these studies were carried out in the money domain and participants made decisions for themselves. As it cannot be assumed that the same effects will show up in other domains, it is important to investigate how real rewards affect risk attitudes in the health domain when participants make treatment decisions for others. For this purpose, half of the subjects in our experiment were allocated to a treatment where their decisions had "real" consequences for patients outside the lab.

3. Experimental Design and Procedure

3.1. *Experimental Design*

The experiment comprises three parts: (i) A MPL task with a monetary outcome, (ii) two MPL tasks with health outcomes, (iii) one MPL task to elicit equity preferences in health.

The results of the two MPL tasks with health outcomes (Part 2 of the experiment) are the main data for our analysis. The MPL task with monetary outcomes (Part 1 of the experiment) was always presented first and functioned as a practice question. The idea behind this is that subjects would more readily understand the MPL task with a familiar outcome like money. The data in Part 3 will be analyzed in a different paper. The order of the tasks of Part 2 (tasks 2h and 2l, for the skewed-high and skewed-low frames) and the equity task of Part 3 were randomized to test for order effects, so that the following sequences were used: 2h-2l-3, 2l-2h-3, 3-2h-2l and 3-2l-2h (note that the two tasks of Part 2 were always adjacent). At the end of the sessions, subjects were asked to fill in a socio-demographic questionnaire.

We used a between-subjects design to respond to our three research questions, and a within-subjects design to test for frame-related biases. In order to implement our design a total of 24 groups (3 health contexts x 2 incentive groups x 4 task sequences) with different questionnaires were needed. Table 1 summarizes our group design. Notice also that each group consisted of a combination of students from health and non-health degrees in a ratio of 2:1.

[Insert Table 1]

3.2. Implementation of the Multiple Price List

The goal of the experiment was to analyze risk aversion across three different medical contexts. In each context, subjects had to choose between pairs of lotteries R and S, where R stands for Risky and S for Safe. Let us assume that four health outcomes are ordered from best to worst as follows: H1>H2>H3>H4. The R lottery will be characterized by (p,H1;H4) and the S lottery by (q,H2;H3) where p and q are the probabilities of the best outcome in each lottery. We used three different types of health outcomes (H), described in subsection 3.3. The “no choice” option (doing nothing) was always stochastically dominated by the R lottery since H4 was at least as good as “no choice.” Following H&L (2002), participants faced a table with 11 decision rows[8] each of which presents a choice between an S gamble (q,H2;H3), labeled treatment A in the experiment, and a R gamble (p,H1;H4), labeled treatment B. In each row p=q. In the first row p=q=0, and in the eleventh row p=q=1. As we move from the first to the eleventh row, the probability of the best outcome increases by 10%. The payoffs or health outcomes remain constant. For this reason, in the first row S stochastically dominates R and in the eleventh row R stochastically dominates S. In every row, the subject has to choose between S and R. As we move down in the table, the expected value of both treatments/lotteries increases as the probability of the high health outcome increases, but the expected value of treatment B increases more quickly than the expected value of treatment A. Thus, only risk-seeking subjects would take Option B in the second decision and only risk-averse subjects would choose Option

[8] H&L have 10 rows, where the last row gives probability 1 to the highest outcome of each lottery. We present 11 rows, adding a row where the lowest outcome of each lottery/treatment get probability 1.

B in the second to last decision. The expectation is that a risk neutral subject should cross over to Option B when the expected value of each option is about the same.

3.3. The Health Contexts

We used three decision contexts corresponding to three types of medical decisions. In each context, the physician does not know the probability of success or failure of the treatment since it changes with unrevealed patient characteristics.

1. Health context 1 (*health gains in years of life in perfect health*): Subjects were told that the patient had been diagnosed with a disease X. They had to choose between two treatments (A or B) that prolong life for several years in perfect health. Outcomes H1 to H4 were described in terms of life years gained in perfect health. We chose this health context because it has been shown that subjects have certain aspirations regarding life expectancy (Verhoef et al., 1994), which can lead to risk-seeking or risk-averse attitudes depending on these aspirations.
2. Health context 2 (*health gains in months of life in moderate ill-health for patients with reduced life expectancy*): Subjects were told that the patient had been diagnosed with a terminal disease X that cannot be cured and that is reasonably expected to result in the death of the patient within a short period. Subjects had to choose between two treatments (A or B) that could prolong life a few months. Outcomes H1 to H4 were described in terms of months of life gained in moderate ill-health.[9] This context is highly relevant as risk and uncertainty are common elements of critical care situations (Nightingale and Grant, 1987). For instance, this type of context is commonly found in cases where critical decision-making leads to the rejection of new cancer drugs because of their high cost.[10]
3. Health context 3 (*health gains in hours of pain alleviated):* Subjects were told that they were treating a patient who had just undergone surgery. The patient would endure acute pain

[9] Following the EQ-5D instrument, we define moderate ill-health as a condition that requires the limitation of daily activities (work, social life, etc.) but presents no problems for personal care (washing, dressing, etc.), does not depend on anyone to move/ and does not cause major pain or anxiety.
[10] http://www.telegraph.co.uk/health/healthnews/11081589/New-drug-for-pancreatic-cancer-turned-down.html

that would be particularly strong in the first 48 hours after surgery but would disappear in 4 or 5 days. Subjects had to choose between treatment A and B in order to alleviate pain. Thus, outcomes H1 to H4 were described in terms of hours of pain alleviated after surgery. Treatment of pain is highly debated within the clinical community, and it is not clear whether more risk-averse physicians are willing to apply stronger painkillers or, on the contrary, they are more concerned about the side effects of painkillers.[11]

The three selected contexts accurately portray the kinds of dilemmas that physicians confront regularly in medical practice. They reproduce a range of conditions from minor to critical (for example, end of life contexts) and thereby establish a basis for comparison between distinct scenarios, which may indicate differences in the degree of risk aversion estimated for each context. We have considered it especially important to include a context of pain treatment and a context of reduced quality of life (end of life context), as it has been observed that preferences for health problems involving quality of life can differ from preferences in situations where outcomes are defined in terms of duration (Attema et al.,2013; Attema, et al. 2016).

Notice that the outcomes H1 to H4 took the same values in each of the three health contexts. What distinguishes the contexts is the type of medical decision faced by the subjects, which implies that the interpretation of outcomes differs across contexts (several years of life-expectancy in perfect health, few months of life-expectancy in moderate ill-health, and hours of pain alleviated). Notice also that none of our three contexts is defined in the loss domain since subjects in the experiment always face choices about possible gains for their patients.

3.4. The frames

In order to control for the potential bias toward selecting the middle row, we followed Andersen et al. (2006) and developed two asymmetric frames in which the implied risk attitudes of the middle row are different. We devised the two frames using a simple procedure inspired by Artinger et al. (2014). In the first frame, H1=39, H2=16, H3=12, and H4=4, and the

[11] See for instance https://www.statnews.com/2017/01/17/chronic-pain-management-opioids/)

interpretation of these figures changes with each health context. So, in the first context, 39 refers to 39 years in perfect health; in the second context it refers to the number of months in moderate ill-health before death; and in the third context it refers to the number of hours of pain alleviated. The second frame we proposed was almost identical to the first except that the best outcome was 19 instead of 39. This use of two frames generated two skewed distributions: Frame "high skew" corresponded to a scenario in which a risk-neutral person would choose the safe option (S) in the first three rows and then would switch to the risky option (R) as indicated by the following pattern SSSRRRRRRR; Frame "low skew", in contrast, corresponded to a situation in which a risk-neutral agent would switch his/her decision from treatment A (safe option) to treatment B (risky option) in row number eight, as indicated by the following pattern SSSSSSSSRRR. Appendix D provides information on how the multiple price list procedure was implemented, first in the case of frame "high skew" and, later, in the case of frame "low skew" for health context 1 (years of life).

3.5. The rewards

All participants in the experiment received a participation fee. The fee amounted to 5 Euros for participants in the "real group" and 15 Euros for those in the "hypothetical group".[12] In the hypothetical group, decisions did not involve additional payments. In the real group, decisions were incentivized and the subjects' choices in the two MPL tasks with health outcomes had "real" consequences for patients outside the lab, which offered an incentive to the participants to take the patient's benefit into account. Subjects in the groups with real payoffs were informed that one of them would be randomly selected at the end of the session and one of his decisions would be used to compute the amount of money transferred to the Vicente Ferrer Foundation. It was explained that the Vicente Ferrer Foundation is an internationally known non-profit institution that works in Annantapur, India, where the institution has built up a network of health services for the poorest populations. As the main aim of this experiment was

[12]On average all participants obtained a payment of 15 euros, since those individuals assigned to the real group received additional money depending on their choices in Part 1 of the experiment.

to analyze whether risk attitudes are context-dependent in the health domain, we designed payments in such a way that differences in subjects' choices could be attributed only to differences of health context and not to changes in the size of the rewards. For this end, we established a monetary equivalence between different health outcomes (years of life gained in perfect health/months of life gained in moderate ill-health/hours of pain alleviated) of 4 euros per time unit.[13] In addition, subjects were made aware that, if selected, they could benefit patients in Annantapur with a payment of up to 156€.

3.6. Experimental Procedure

The experiment was programmed with z-tree (Fischbacher, 2007) and conducted at the Laboratory for Research in Experimental Economics (LINEEX) in May 2014.[14] The laboratory recruited a total of 257 undergraduate students from the University of Valencia following standard electronic recruitment procedures. None of the subjects had participated in a similar economic experiment before.

We conducted 5 sessions of approximately 45 minutes at the lab to collect the data. Students were allocated to one of the 24 potential treatments using quotas so as to have (approximately) the same number of observations for each treatment. Within each treatment a 2:1 ratio was maintained between health and non-health students. Sessions 1, 2 and 3 dealt with health contexts 1, 2 and 3 respectively. Within each group of students, each computer was assigned one of the 8 potential treatments (2 rewards groups x 4 task sequences). In sessions 4 and 5 all health contexts were used. In this case, students were allocated to one of the 24 potential treatments until the quota was full. Table 2 below summarizes the allocation of the participants in health context 1 (n=85). A similar allocation was made in the other two health contexts (n=86 each).

[13] An interesting extension of our study would be to investigate whether individuals' risk attitudes within the same health context might vary according to the life horizon faced by their patients. Physicians, for instance, may be more or less risk averse within an end-of-life context depending on whether the magnitude of patients' remaining life expectancy is shorter or longer (weeks instead of a few months, or days instead of a few weeks, for instance). Investigating this issue would likely require a recalibration of payments in the real group in order to measure all the outcomes in the same unit. We leave this question open for future research.

[14] Previously we performed a pre-test with 45 students of economics at the Public University of Navarra to confirm that the experiment instructions were clear and the answers did not present any major problems.

[Insert Table 2]

The experimental procedure in all sessions was as follows: upon arrival, subjects were randomly assigned to cubicles, where they responded to the questions in full anonymity. The general instructions of the experiment were read aloud. Once the experiment started, specific instructions were shown to participants on the screens of their computers.

Subjects were privately paid at the end of the experiment. In addition, a total of 341 euros was transferred to the Vicente Ferrer Foundation. This amount reflects the money earned in the experiment as a result of the medical choices of the selected participants in the groups with real payoffs. To validate the actual transfer of money to the foundation, an email was sent to all participants after the experiment with a scanned copy of the check showing the total amount collected. Experimental instructions are provided in Appendices B, C, D and E.

4. Methods

We assume that risk preferences can be represented by a constant relative risk aversion (CRRA) utility function over the health outcomes (*H*), with a coefficient of constant relative risk aversion *r*:[15]

$$U\left(H;r\right)=\frac{H^{1-r}}{1-r} \qquad (1)$$

We can infer r based on the 11 decisions made by the individual in each of the two frames analyzed. When individuals switch from the safe option (treatment A) to the risky option (treatment B), an interval is elicited where the risk aversion parameter r is located.

As in H&L (2002), we can find the lower (r^{lb}) and upper bound (r^{ub}) of the interval of r associated to the switching point k'. If p^k denotes the probability of the best health outcome in each of the decision rows k, with $k=\{1,...11\}$, then we can define r^{lb} and r^{ub} as:

$$r^{lb}=\left\{r: p^{k'}U\left(H2;r\right)+\left(1-p^{k'}\right)U\left(H3;r\right)=p^{k'}U\left(H1;r\right)+\left(1-p^{k'}\right)U\left(H4;r\right)\right\}, \text{ and}$$

[15] Depending on the value of *r*, individuals exhibit different degrees of risk aversion in the health domain that can be grouped into three types: risk neutral (if *r*=0), risk averse (if r>0) and risk seeker (if r<0).

$$r^{ub} = \left\{ r : p^{k'+1} U\left(H2; r\right) + \left(1 - p^{k'+1}\right) U\left(H3; r\right) = p^{k'+1} U\left(H1; r\right) + \left(1 - p^{k'+1}\right) U\left(H4; r\right) \right\},$$

Table 3 reports the implicit r intervals for the two skewed frames described in the experimental design section. For subjects with multiple switching points we follow Andersen et al. (2006). Thus, when individuals go back and forth between safe and risky options as they move along the decision rows, we assume they are indifferent to the options in the intermediate decision rows, and therefore we define r^{lb} using the first switch and r^{ub} using the last switch. For example, someone might have the following preferences: SSSRSRRRRRR, that is, three safe choices, then risk in the fourth row, back to safe in the fifth row and, finally, switch to risk from sixth to 11th. In this case r^{lb}=-0.27 and r^{ub} =0.83 for Frame "high skew".

[Insert Table 3]

The coefficient of relative risk aversion of individual *i* (r_i) can be assumed to be a linear function of individual characteristics (X_i) and a stochastic term (e_i), as defined in equation (2).

$$r_i = X_i \beta + e_i \qquad (2)$$

Individual characteristics X_i include whether the subject is enrolled in a health related degree, age, gender, self-reported health status, weekly disposable income and level of education of his/her parents, among others.

Since the only observable data is the interval $\left[r_i^{lb}, r_i^{ub}\right]$, inferences about β and predictions of r_i can be obtained using an interval regression model as described by Harrison et al. (2005). In particular, $\Pr\left(r_i^{lb} \le r_i \le r_i^{ub}\right) = \Pr\left(r_i^{lb} - X_i\beta \le e_i \le r_i^{ub} - X_i\beta\right)$, which is used to estimate the model via maximum likelihood. The error term is assumed to follow a normal distribution with censoring, which accounts for choices in the extremes of the decision rows when r^{lb} or r^{ub} goes to infinity. The β parameter related to the dummy variable "health related major" is used to test the effect of the type of degree on the coefficient of risk aversion (research question 2 described in Section 2).

In order to answer research questions 1 and 3 (see Section 2) we add a set of dummy variables E_i reflecting the design of the experiment. That is, we use model (3):

$$r_i = X_i\beta + E_i\theta + e_i \quad (3)$$

Three sets of dummy variables are used. First, to test the effect of monetary incentives we include a dummy variable (E_1) where $E_1 = 0$ for the hypothetical group and $E_1 = 1$ for the group with financial incentives. The null hypothesis, $\theta = 0$, indicates that financial incentives do not influence the magnitude of the coefficient of relative risk aversion. The second set (E_2) corresponds to the health context (health gains in years of life, months of life, or hours of avoided pain). Finally, another set of dummies (E_3) indicates the task sequence assigned to the subject, which is used to test for order effects.

We test for frame-related biases (what we call "skewness effects") by exploiting the within-subjects design of the experiment. As in Andersen et al. (2006), we use the two skewed frames faced by all individuals to form a panel of observations. Let r_{it} be the coefficient of relative risk aversion of individual *i (i=1,…,N)* inferred from frame *t (t=H,L)*. Equation (3) can now be rewritten as:

$$r_{it} = X_i\beta + E_i\theta + T_{iH}\gamma + v_i + e_{it} \quad (4)$$

Where $T_{iH} = 1$ if individual i performs frame "high skew", and 0 otherwise. v_i is an individual random effect that captures unobserved individual variables and allows for the correlation of responses from the same subject (Harrison and Ruström, 2008b). The random effect is, by assumption, uncorrelated to observed variables X, E and T. Equation (4) is estimated via panel data interval regression model. The skewness effect is tested by inference of parameter γ. The null hypothesis, $\gamma = 0$, suggests that the skewed-high or low frames do not affect the magnitude of the coefficient of relative risk aversion, thus indicating no skewness effect.

5. Data and Results

5.1. Characteristics of individuals

Table 4 reports descriptive statistics of all individuals participating in the experiment. The average age of subjects was 23; 61% were female and 69% were students from medicine and other health-related majors. The two columns of table 4 show the characteristics of the participants separated by health-related and non-health related majors.[16] There were slightly more females in health-related vs. non-related majors (69% versus 47%), and most subjects were in their fourth year of studies. The table also includes individual characteristics (e.g. age, gender, labor activity, etc.), characteristics of the experimental design as described in Section 3, and the proportion of multiple switches by frame, which was similar between health and non-health related students (9% versus 10%).

[Insert Table 4]

5.2.Data Quality

The data consists of 514 observations (257 subjects performing two task frames). Table 5 shows the distribution of safe choices according to contexts and frames. There is clearly a change in the pattern of responses from one frame to the other as rationality dictates, that is, when the best outcome of the R lottery is 19, subjects chose the S lottery more often than when it is 39. The proportion of inconsistent choices is 9.3% of the total sample (N=48). It includes choices by subjects who switched several times (5.6%, N=29), choices by non-switchers (1.7%, N=9) and choices by those who select the risky choice in the first decision row (1.9%, N=10).[17] Notice that the proportion of multiple switchers is small compared to the range of 6% to 21% found in previous studies performing the H&L test (H&L, 2002; Lusk and Coble, 2005; Anderson and Mellor, 2008).[18]

[Insert Table 5]

5.3. *Estimation of Coefficient of Relative Risk Aversion and results*

16 Health-related majors include students of medicine (n=62, or 35% of health-related students), nursing (n=39, 22%), physiotherapy (n=35, 20%), pharmacy (n=14, 8%), and odontologists (n=28, 16%).

17 Multiple switchers and non-switchers that select the risky choice in the first decision row are included in the 1.9%.

18 According to the recent survey by Filippin and Crosetto (2016) the proportion of multiple switchers is about 10% in experiments using the H&L elicitation method.

Table 6 reports the elicited coefficients of relative risk aversion as a result of an interval regression where we control for individual and experimental design characteristics.[19] The table presents individual regressions for the skewed-high and the skewed-low frames, as well as an unbalanced panel data regression that combines both frames, assuming a random effect. In general, the estimates of the combined frames (column c) fall within the interval defined by the estimates of the skewed-high (column a) and skewed-low (column b) frames.[20]

Of the 514 observations, nine (1.7% of the sample) were excluded from the estimation. These correspond to choices by those who selected the risky choice in the first decision row and then either switched to safe choices or continued with risky choices in the following decision rows, which prevents us from obtaining their implicit coefficient of relative risk aversion.[21] The unbalanced panel data includes a total of 255 subjects and 505 observations.[22]

Table 6 shows that the health context where individuals decide on hours of pain alleviated for their patients is associated with higher risk aversion. This result is statistically significant for the skewed-high frame (column a) and for the estimation that combines both frames (column c), where the coefficient of relative risk aversion is higher in context 3 compared to context 1 by a margin of 0.348 points. No statistical differences are found between the other two contexts. **This result responds to our first research question.**

[Insert Table 6]

Individual characteristics barely affect the coefficient of relative risk aversion. However, one important result that allows us to **respond to research question 2** is that being enrolled in a medical degree or other health field increases the coefficient of relative risk aversion by 0.308

19 We also tested for interaction effects between contexts and hypothetical and real groups and between these groups and skewness effects. We did not find statistically significant interactions effects and the average coefficient of relative risk aversion estimated for the model with and without interaction effects does not show statistically significant differences either (1.5635 versus 1.5632).

[20] The interval regression model was also estimated in a stepwise manner: first, without controls, then, adding controls one by one. Results obtained in the model with all the variables and controls included are mostly similar to the ones obtained from the regression without controls and those obtained in the intermediate regressions.

21 Two subjects showed this anomalous behavior in both frames and as result their corresponding four observations were excluded from the sample. Another five subjects made this anomalous choice only in one of the two frames, so just their five corresponding observations were excluded. Notice that we included, however, one observation where the risky choice is chosen in the first decision row but then there are multiple switches. In this case, we estimated the coefficient of relative risk aversion by considering the first row where the subject switched from a safe to a risky decision.

22 If we exclude all multiple switchers and non-switchers from our sample, the results hold and we do not find statistically significant differences between the average coefficient of relative risk aversion in the two estimations (1.5635 versus 1.5418).

points (Table 6, column c), i.e. students in health-related fields seem to be more risk averse than those in non-health-related fields. However, this result is not observed in the skewed-high frame, where the difference is negligible.

Table 6 also indicates that, when real payments are used, the coefficient of relative risk aversion is lowered by 0.377 points (column c), which **addresses research question 3**. This result is observed in the individual regressions by frames and also in the regression that combines both frames. Our results also show a non-statistically significant order effect.

5.4. Skewness effects

In Table 6 we find a statistically significant skewness effect. The implicit coefficient of relative risk aversion is 0.384 points lower in the skewed-low frame compared to the skewed-high frame (column c). To further explore the skewness effect, we present the kernel distribution of predicted r_i by skewed frame. Predicted r_i are based on interval regression estimation of equation (3) for each frame, whose parameter estimates are also presented in columns a and b of Table 6. The average coefficients of relative risk aversion are estimated to be 1.72 and 1.43 when based on skewed-high frame and skewed-low frame, respectively; when both frames are combined in the panel data interval regression the average coefficient of relative risk aversion is 1.56. Figure 1 shows that the distribution of r_i in the skewed-high frame is not only located to the right, avoiding negative values of r, but it is also less disperse, with values more concentrated around the mean.

[Insert Figure 1]

5.5. Discussion

Our analysis of the risk preferences across different types of medical decisions made on behalf of a patient yields several interesting results for discussion.

First, we find that preferences show a high degree of risk aversion in all health contexts, with risk-averse preferences in nearly 100% of the subjects. This is a new and potentially significant result because almost every other paper that has measured risk aversion in health has shown a mixture of attitudes toward risk (VdP&R, 2012; Schosser et al 2016). The result is also outstanding because it was obtained even when one of the frames, the skewed-low frame, invited subjects to show risk-seeking preferences. The divergence of results could also be related to the fact that we focused on decisions for others, while previous works did not. Thus, our prevalence of risk aversion would be in line with the findings in Ubel et al (2011) and Garcia-Retamero and Galesic (2012), namely that doctors more often select a safer medical treatment for their patients than for themselves.

Second, we find differences between medical contexts since individuals are even more risk averse in decisions involving hours of pain alleviated than in the other two contexts. This result is surprising, as one might expect that if the number of QALYs involved is smaller, as it is in this context, risk aversion should be also smaller (as happens with monetary outcomes). We wonder whether our result is driven by the fact that participants in the experiment saw an outcome of hours of pain as more plausible and real than an outcome involving shortened life expectancy, and were therefore more cautious in the former context. Whatever the explanation, our result seems to indicate that some of the economic theory predictions on attitudes towards risk with monetary outcomes may have limited validity when applied to health outcomes. Further, the fact that our results indicate that subjects treat pain differently to other health outcomes (at least when making decisions for others), calls into question the adequacy of the QALY model when it is used to represent preferences of outcomes involving pain (very common in health problems). This is a significant finding, given that the QALY model is the dominant method for the economic evaluation of medical treatments. In fact, Mehrez and Gafni (1989) used examples with pain to show the limitations of the QALY model, suggesting, instead, an alternative such as the Healthy Year Equivalent. More recently, Schosser et al. (2016) showed that preferences for duration of pain are different than preferences for intensity of pain. Our findings are aligned with this evidence and suggest that the same utility function should not be used for all kinds of health outcomes. The utility function for health, even if it is a QALY-type utility function, should use different parameters for different health outcomes.

Thirdly, the fact that students in health-related degrees seem to be more risk averse than those of other degrees implies that doctors can be highly risk-averse within the context of their medical practice. There are two potential explanations for this finding: the self-selection of risk-averse students in health-related degrees and/or the possibility that health-related students have a better understanding of the health decisions they confront. In fact, because participants in the experiment answered H&L questionnaires in both the monetary domain and the health domain, we were able to compare these questionnaires and found that health-related students are more risk-averse than non-health related students in the health domain but less risk-averse in the monetary domain (see Appendix A for these results).[23] This result calls into question the linearity of the QALY model, which assumes risk neutrality (Loomes and Mckenzie, 1989). Other papers have already questioned the linearity assumption of the QALY model when subject make decisions for themselves (e.g., Bleichrodt and Pinto, 2005; Schosser et al., 2016); we question the linearity assumption when subjects make medical decisions as agents for others.

Fourth, our results show a lower degree of risk aversion in the group with real payoffs. This result contrasts with previous findings in the money domain (H&L, 2002, 2005 and Harrison and Rutström, 2008b), where opposite results are found. We can think that real rewards caused subjects to take the experimental task more seriously, so that the change in risk attitudes is the consequence of more careful deliberation about the health consequences for patients. However, we do not find that real pay-offs led to fewer inconsistencies. Also, subjects in the hypothetical group took longer to respond to the MPL tasks (178 SEC VS 140 SEC, P<0.01), which seems to confirm that they took at least the same interest in responding than the subjects with real payoffs. An alternative explanation may be that our participants in the experiment saw a real patient in India as too distant, affecting the credibility of the rewards and/or decreasing the perceived size of the rewards and leading, as a result, to less risk

[23]We need to be cautious when interpreting this comparison between H&L in the monetary domain and the H&L in the health domain, since the first one was designed just as a practice exercise for participants and as a result participants were playing for themselves (and not for others as in the H&L in the health domain) and the lotteries were not skewed as they were for the health domain.

aversion. Recent findings in the literature indicate that there is a negative correlation between risk aversion and aversion to inequality or inclination towards social preferences in a context of giving donations or giving to others (Müller and Rau 2016). This might explain the fact that individuals are willing to take more risk (i.e. are less risk-averse) when real money can be transferred to a NGO, as opposed to a hypothetical situation where no real money is involved. In any case, whatever the reason, the effect (change in the coefficient of risk aversion) was not very large. Other studies find no differences between hypothetical groups and groups with real payoffs as mentioned in the introduction. Thus, this seems to be an issue that needs further exploration.

Finally, the two skewed frames produced different degrees of risk aversion. The vast majority of subjects had a higher degree of risk aversion in the skewed-high frame, which shows that skewness effects did occur. However, the effects were not very important: on aggregate the difference between the two means (1.72 vs. 1.45) does not seem notable. Even with skewness effects, we can definitely say that there is evidence of risk aversion in the health domain.

Several limitations of this study deserve mention here. One is related to the effect of monetary incentives. When answering the H&L in the health domain with real pay-offs, it is possible that subjects first convert the health outcomes into money, while subjects in the hypothetical group focus simply on the health outcomes. In this case, the difference between real and hypothetical groups might be the result of differential risk attitudes between health and monetary outcomes, and not due to the incentives. To ascertain the likelihood of this conversion, we have compared answers to H&L questionnaires in the monetary and health domains, observing a higher average risk-aversion coefficient in the health domain (see Appendix A). In both cases, there are differences between the real and hypothetical groups. However, in the monetary domain subjects are more risk averse in the group with real pay-offs as compared to the hypothetical group, while the opposite occurs in the health domain. Given this result, we cannot rule out the possibility that the group with real pay-offs in the health domain is monetizing the health outcomes.[24] This result shows that introducing the right monetary

[24] Again, comparisons between the monetary and the health domains should be made with caution. See footnote 24.

incentives when dealing with health decisions is a challenge that needs further exploration. Another key question of experimental design was to establish the marginal rate of substitution between health and money. We established a monetary equivalence of 4 euros per time unit (years/months/hours) for each of the three different health contexts. This rate of exchange was chosen so that the total amount that could be transferred to the NGO would be perceived as substantial by the participants. However, Galizzi et al. (2016b) followed a different approach that established one day of full health as equivalent to one monetary unit. Establishing this kind of equivalence is not a trivial exercise and it should be noted that different approaches can lead to different results.

Finally, we wish to acknowledge two methodological limitations. The first is related to our use of the Expected Utility framework, which prevents us from considering probability transformations in our model (Kahneman and Tversky, 1979). Thus, it is possible that some of the risk aversion we observed could be attributed to probability transformations. However, it is also true that preference representations that employ probability transformations are mainly relevant when the probability of the event is near to the end-points 0 and 1 (there is no behavioral bias when events are certain or impossible), and in this paper we do not consider probabilities of events near to 0 and 1. The second limitation is related to estimation strategy: our panel data interval regression approach relies on the assumption of a CRRA utility function and does not model explicitly behavioral errors. However, the literature shows that our approach produces relatively consistent estimates of CRRA coefficients (Harrison and Rutström, 2008b). Additional approaches such as structural estimation methods could complement and enrich this type of approach, but these are more commonly used with more flexible utility functions or when utility functions include more than one parameter (see, e.g., Anderson et al., 2008), which is not our case.

6. *Conclusion and future research*

Although it is widely recognized that individual risk attitudes are an important factor in the health domain, risk attitudes in medical decisions for others have hardly been studied, and elicitation techniques that are widely used in economics—such as the multiple price list—have not been applied systematically.

In this paper we carried out a lab experiment that investigated the risk attitudes of a group of 257 individuals who were assigned the role of a physician deciding between treatments for a patient. Our main goal was to elicit their risk preferences for others and test whether this individual risk attitude is context dependent. We used three different decision contexts with health gains expressed in (i) years of life in perfect health, (ii) months of life in moderate-ill health for patients with reduced life expectancy, and (iii) hours of pain alleviated, and elicited risk preferences for each context using a multiple price list method.

Results of our study shed light on issues hardly explored until now and also raise interesting questions for future research. First, our results indicate that when decisions are made on behalf of a third party, the introduction of real rewards seems to decrease risk aversion. In our experiment, participants' decisions determined the monetary rewards that were to be transferred to the Vicente Ferrer Foundation. It would be interesting to explore alternative ways to incentivize medical decisions on behalf of a third party, as well as analyze how different incentives for physicians influence their risk preferences and treatment choices when making decisions for others. Secondly, future research should also explore why our results find such a large prevalence of risk aversion, unlike previous findings in the field of health. A comprehensive study that uses the same preference elicitation method to compare medical decisions for self-treatment versus treatment for others within different health contexts would help to clarify some of the issues and questions raised in the discussion. Finally, it is important to question the extent to which responses to a task like the MPL can predict risk preferences in real-life decisions. As mentioned in the Introduction, the H&L method overcomes some of the disadvantages of standard gamble methods, which are the most commonly used methods to elicit preferences towards health risks. At the same time, it is a feasible and convenient method that typical subjects (i.e. university students) can use without much confusion. But whether doctors are really so risk averse in practice is a question that needs further exploration. Galizzi et al., 2016b conducted a frame field experiment that found mild but significant risk aversion among doctors in the health domain, and a more pronounced risk aversion in the monetary domain. Moreover, the question of whether risk preferences that are elicited in laboratories using small populations and relatively low stakes can be transferred to medical contexts is a contentious issue in the experimental economics literature (Charness et al., 2013) and reflects a wider debate concerning the validity of lab experiments for explaining behaviors in natural settings (see Kessler and Vesterlund, 2015; Gallizzi and Wiesen, in press). As Galizzi and Wiesen (in press) remark, behavioral and experimental health economists must confront the challenge of establishing the external validity and generalizability of laboratory results. Indeed, the comparison of experimental results with data from the field is an important area for future investigation.

References

Abdellaoui, M., Driouchi, A., and L'Haridon, O. (2011). Risk aversion elicitation: Reconciling tractability and bias minimization. *Theory and Decision* 71(1), 63-80.

Abdellaoui, M., Bleichrodt, H., Kammoun, H. (2013): Do Financial Professionals Behave According to Prospect Theory? An Experimental Study, in: *Theory and Decision* 74, S. 411–429.

Andersen, S., Harrison, G.W., Lau, M.I., Rutström, E. (2008). Eliciting risk and time preferences. *Econometrica* 76 (3), 583–618.

Andersen, S., Harrison, G.W., Igel Lau, M. and Rutström, E. (2006). Elicitation using multiple price list formats. *Experimental Economics* 9, 383-405.

Andersson, O., Hakan, J.H., Tyran, J-R. and Wengstrom, E. (2016). Deciding for Others Reduces Loss Aversion. *Management Science* 62(1): 29-36.

Anderson, L.R. and Mellor, J.M. (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of Health Economics* 27, 1260-1274.

Artinger, F., Brown, G. D. A., Loomes, G.C. (2014). Risk Preferences in Context. Mimeo.

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

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

Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C. and Sugden, R. (2010). Experimental Economics: Rethinking the Rules. Princeton University Press, Princeton, NJ.

Barseghyan, L., Prince, J., Teitelbaum, J. (2011). Are risk preferences stable across contexts? Evidence from insurance data. *American Economic Review* 101 (2), 591-631.

Bleichrodt, H., and Pinto, J. L. (2005). The validity of QALYs under non-expected utility. *The Economic Journal 115*(503), 533-55.

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

Brosig-Koch, J., Hennig-Schmidt, H., Kairies, N., Wiesen, D., 2016. Using artefactual field and lab experiments to investigate how fee-for-service and capitation affect medical service provision. *Journal of Economic Behavior and Organization* 131B:17-23.

Buckley, N.J., Cuff, K., Hurley, J., McLeod, L., Mestelman, S., Cameron, D. (2012). An experimental investigation of mixed systems of public and private health care finance. *Journal of Economic Behavior Organization* 84, 713-729.

Butler, S., Rosman, A., Seleski, S., Garcia, M., Lee, S., Barnes, J., & Schwartz, A. (2012). A medical risk attitude subscale for DOSPERT. *Judgment and Decision Making*, 7, 189–195.

Byrnes, James P., Miller, David C., and Schafer, William D. (1999). Gender Differences in Risk Taking: A Meta-Analysis. *Psychological Bulletin*, 125: 367-383.

Camerer, C.F. and Hogarth, R.M. (1999). The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty*, 19 (1-3), 7–42

Chakravarty, S., Harrison, G., Haruvy, E. and Rutström, E. (2011). Are You Risk Averse over Other People's Money? *Southern Economic Journal* 77(4), 901–13.

Chandra, A., Cutler, D. and Song, Z. (2012). Who ordered that? The Economics of Treatment Choices in Medical Care. Chapter of Handbook of Health Economics Vo.2. Elsevier B.V.

Charness, G., Gneezy, U., and Imas, A. (2013). Experimental methods: Eliciting risk preferences, *Journal of Economic Behavior and Organization* 87, 43-51.

Charness, G., Viceisza, A., (2017). Comprehension and risk elicitation in the field: Evidence from rural Senegal. Review of Behavioral Economics. Forthcoming.

Cohen, J.D., Ericson, K.M., Laibson, D. and White, J.M. (2016). Measuring Time Preferences. NBER Working Paper No. w22455. Available at SSRN: https://ssrn.com/abstract=2816729.

Dohmen, T., Falk, A. (2011). Performance pay and multidimensional sorting: productivity, preferences and gender. *American Economic Review* 101(2): 556-590

Dyer, D., Kagel J.H. and Levin, D. (1989). A Comparison of Naive and Experienced Bidders in Common Value Offer Auctions: A Laboratory Analysis. *Economic Journal*, 99: 394 (March), 108-115.

Einav, L., Finkelstein, A., Pascu, I., Cullen, M.R., (2012). How general are risk preferences? Choices under uncertainty in different domains. *American Economic Review*, 102(6), 2606-2638.

Eriksen, K.W. and Kvaloy, O. (2010). Myopic Investment Management. *Review of Finance* 14(3): 521-542.

Farquhar, P.H. (1984). State of the art—utility assessment methods. *Management Science*, 30(11), 1283-1300.

Filippin, A., Crosetto, P. (2016). A Reconsideration of Gender Differences in Risk Attitudes. Management Science, 62 (11): 3138-3160.

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-made Economic Experiments, *Experimental Economics* 10(2), 171-178.

Galizzi, M.M., Miraldo, M. and Stavropoulou, C. (2016a). In Sickness but Not in Wealth: Field Evidence on Patients' Risk Preferences in Financial and Health Domains. *Medical Decision Making* 36, 503-517.

Galizzi, M.M., Miraldo, M., Stavropoulou, C., Van Der Pol, M. (2016b). Doctors-Patients Differences in Risk and Time Preferences: a Field Experiment, Journal of Health Economics, 50, 171-182.

Galizzi, M.M., Wiesen, D. (forthcoming). Behavioural Experiments in Health Economics. In: Oxford Encyclopedia of Economics and Finance. Oxford University Press.

Garcia-Retamero, R., and Galesic, M. (2012). Doc, what would you do if you were me? On self-other discrepancies in medical decision making. *Journal of Experimental Psychology Applied* 18(1), 38-51.

Godager, G., Wiesen, D., (2013). Profit or patients' health benefit? Exploring the heterogeneity in physician altruism. *Journal of Health Economics* 32, 1105-1116.

Hanoch, Y., Johnson, J.G., & Wilke, A. (2006). Domain specificity in experimental measures and participant recruitment. Psychological Science, 17, 300-304.

Harrison, G.W., Lau, M.I., Rutström, E.E., Sullivan, M.B., (2005). "Eliciting Risk and Time Preferences Using Field Experiments: Some Methodological Issues". In: Field Experiments in Economics, Research in Experimental Economics, Vol. 10, ed. by J. Carpenter, G.W. Harrison, and J. A. List. Greenwich CT: JAI Press, 125–218. [439,445].

Harrison, G.W. (2006). Hypothetical bias over uncertain outcomes. In: Using Experimental Methods in Environmental and Resource Economics. Eds: John A. List, Northampton, MA: Edward Elgar.

Harrison, G.W., List, J.A., and Towe, C. (2007a). Naturally occurring preferences and exogenous laboratory experiments: a case study of risk aversion. *Econometrica* 75 (2):433-458.

Harrison, G.W., Lau, M.I. and Rutström, E.E. (2007b). Estimating risk attitudes in Denmark: A field experiment. Scandinavian Journal of Economics 109(2):341–368.

Harrison, G.W. and Rutström, E.E. (2008a). Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods. Handbook of Experimental Economics Results, Elsevier.

Harrison, G.W. and Rutström, E.E. (2008b). "Risk Aversion in the Laboratory". In: Risk Aversion in Experiments, Research in Experimental Economics, Volume 12, ed. by J.C. Cox and G.W. Harrison. Bingley, UK: Emerald, 41-196.

Harrison, G.W., Lau, M.I and Rutström, E.E (2010). Individual discount rates and smoking: Evidence from a field experiment in Denmark. Journal of Health Economics 29, 708-717.

Hennig-Smith, H., and D. Wiesen (2014). Other-regarding behavior and motivation in health care provision: An experiment with medical and non-medical students. *Social Science and Medicine* 108, 156– 165.

Hertwig, R. and Ortmann, A. (2001). Experimental practices in economics: a methodological challenge for psychologists? *Behavioral and Brain Sciences* 24(3): 383-403.

Holt, C.A and Laury, S.K. (2002). Risk aversion and incentive effects. *American Economic Review* 92(5), 1644-1655.

Holt, C.A. and Laury, S.K. (2005). Risk aversion and incentive effects: New data without order effects. *American Economic Review*, 95(3), 902-12.

Holt, C.A. and Laury, S.K. (2013). Assessment and Estimation of risk preferences. Handbook of Economics of Risk and Uncertainty. Eds: Machina, M and Viscusi, V. North-Holland.

Jagsi R, Lehmann L.S. (2004). The Ethics of Medical Education. *British Medical Journal*, 329 (7461): 332-4.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2): 263-291.

Kessler, J.B. and Vesterlund, L. (2015). "The External Validity of Laboratory Experiments: The Misleading Emphasis on Quantitative Effects". In: Guillaume R. Fréchette and Andrew Schotter, editors, Handbook of Experimental Economic Methodology. Oxford University Press.

Lévy-Garboua, L., Maafi, H., Masclet, D. and Terracol, A. (2012). Risk aversion and framing effects. *Experimental Economics* 15, 128-144.

Loomes, G., and McKenzie, L. (1989). The use of QALYs in health care decision making. *Social Science and Medicine* 28(4), 299-308.

Lusk, J.L., and Coble, K.H. (2005). Risk perceptions, risk preference, and acceptance of risky food. *American Journal of Agricultural Economics* 87(2), 393–405.

MacCrimmon, K.R., and Wehrung, D.A., (1990). Characteristics of risk taking executives. *Management Science* 36, 422–435.

Massin, S., Nebout, A., Pulcini, C., Ventelou, B. and Verger, P. (2015). Cross-sectional survey: risk-averse French general practitioners are more favorable toward influenza vaccination. *Vaccine* 33 (5): 610-614.

Michel-Lepage, A., Ventelou, B., Nebout, A., Verger, P. and Pulcini, C. (2013). Cross-sectional survey: risk-averse French GPs use more rapid-antigen diagnostic tests in tonsillitis in children. *British Medical Journal* Open 3: 10.

Mehrez, A. and Gafni, A. (1989). Quality-adjusted life years, utility theory, and healthy-years equivalents. *Medical Decision Making* 9 (2): 142-149.

Müller, S. and Rau, H.A. (2016). The relation of risk attitudes and other-regarding preferences. A within-subject analysis. *European Economic Review* 85: 1-7.

Nightingale, S.D., and Grant, M. (1987). Risk Preference and Decision Making in Critical Care Situations. *CHEST* Journal 93 (4): 684–687.

Reynolds, D.B., Joseph, J. and Sherwood, R. (2009). Risky shift versus cautious shift: Determining differences in risk taking between private and public management decision-making. *Journal of Business & Economics Research* 7(1): 63-78.

Ruggeri, M. and van der Pol, M. (2012). Is risk attitude really specific within the health context domain? Further evidence from an Italian survey using probability equivalent technique and face-to-face interviews. *Health, Risk and Society* 14 (7-8), 655-666.

Schram, A., Sonnemans, J., (2011). How individuals choose health insurance: an experimental analysis. *European Economic Review* 55, 799-819.

Schosser, S., Trarbach, J. and Vogt, B. (2016). How does the perception of pain determine the selection between different treatments? *Journal of Ecomomic Behavior & Organization*, 131 (PB), 174-182.

Sommers, B.D. and Zeckhauser, R. (2008). Probabilities and preferences: What economics can teach doctors and patients making difficult treatment decisions. *Urologic Oncology* 26(6), 669-673.

Sutter, M. (2009). Individual Behaviour and Group Membership: Comment. *American Economic Review* 99(5): 2247-2257.

Ubel, P.A., Angott, A.M., and Zikmund-Fisher B.J. (2011). Physicians Recommend Different Treatments for Patients than they Would Choose for Themselves. *Archives of Internal Medicine*, 171(7), 630-634.

Van der Pol, M. and Ruggery, M. (2008). Is risk attitude outcome specific within the health domain? *Journal of Health Economics* 27, 706-717.

Verhoef, L.C.G., De Haan, A. F. J., and Van Daal, W.A.J. (1994). Risk attitude in gambles with years of life: Empirical support for prospect theory. *Medical Decision Making* 14(2): 194-200.

Weber, E.U., Blais, A.-R., Betz, E. (2002). A Domain specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, 15, 263–290.

Table 1: Summary of Groups Design.

HEALTH CONTEXTS	Rewards	Task Order				Total subgroups
Health Context 1 (years of life in perfect health)	G1 HYPOTHETICAL	2h-2l-3	2l-2h-3	3-2h-2l	3-2l-2h	4
	G1 REAL	2h-2l-3	2l-2h-3	3-2h-2l	3-2l-2h	4
Health Context 2 (months of life in moderate ill-health)	G2 HYPOTHETICAL	2h-2l-3	2l-2h-3	3-2h-2l	3-2l-2h	4
	G2 REAL	2h-2l-3	2l-2h-3	3-2h-2l	3-2l-2h	4
Health Context 3 (hours of pain alleviated)	G2HYPOTHETICAL	2h-2l-3	2l-2h-3	3-2h-2l	3-2l-2h	4
	G2 REAL	2h-2l-3	2l-2h-3	3-2h-2l	3-2l-2h	4
						24

Table 2: Allocation of Subjects in Health Context 1

Health Context 1 (years in perfect health) N= 85	Real payoffs N=42		Hypothetical Payoffs N=43		
Task sequence (order)	Health Related Subjects	Non Health Related Subjects	Health Related Subjects	Non Health Related Subjects	Total
Skewed-high first, skewed-low second	8	3	8	3	22
Skewed-high second, skewed-low first	7	4	7	4	22
Skewed-high second, skewed-low third	7	4	7	4	22
Skewed-high third, skewed-low second	7	2	7	3	19
Total	29	13	29	14	85

Table 3. Implicit intervals of relative risk aversion by frame

Number of safe choices	Interval of relative risk aversion	
	Frame "high skew"	Frame "low skew"
1	-∞<r≤-0.93	-∞<r<=-4.79
2	-0.93<r≤-0.27	-4.79<r≤-3.39
3	-0.27<r≤0.16	-3.39<r≤-2.52
4	0.16<r≤0.51	-2.52<r≤-1.85
5	0.51<r≤0.83	-1.85<r≤-1.28
6	0.83<r≤1.15	-1.28<r≤-0.73
7	1.15<r≤1.49	-0.73<r≤-0.17
8	1.49<r≤1.9	-0.17<r≤0.47
9	1.9<r≤2.52	0.47<r≤1.36
10	2.52<r≤3.37	1.36<r≤2.61
11	3.37<r≤∞	2.61<r≤∞

Table 4. Individual characteristics and Experimental Design

Characteristics	*Non-health related (n=79)*	*Health Related (n=178)*
Individual		
Age	23.53 (5.51)	23.43 (7.39)
Gender (0=male, 1=female)	0.47 (0.50)	0.67 (0.47)
Health status (1=Excellent, 2=Very good, 3=Good, 4=Regular, 5=Poor)	2.18 (0.80)	1.99 (0.76)
Labor activity last week, none	0.72 (0.45)	0.79 (0.41)
Labor activity last week, part time (less than 15 hours)	0.18 (0.38)	0.13 (0.34)
Labor activity last week, full time (more than 15 hours)	0.10 (0.30)	0.08 (0.28)
Weekly disposable income (in Euros)	39.96 (62.72)	56.38 (74.47)
Year in college (1st to 7th year)	3.57 (2.00)	3.64 (2.26)
Grade of access to college (scale 0 to 1)	0.68 (0.13)	0.76 (0.10)
Number of people per room in household	0.56 (0.21)	0.56 (0.28)
Education of father (0=none, 1=school, 2=college, 3=post-graduate)	1.47 (0.71)	1.74 (0.75)
Education of mother (0=none, 1=school, 2=college, 3=post-graduate)	1.44 (0.66)	1.74 (0.62)
Experimental design		
Order 1 (skewed-high first, skewed-low second), fraction	0.23 (0.42)	0.27 (0.45)
Order 2 (skewed-high second, skewed-low first), fraction	0.30 (0.46)	0.24 (0.43)
Order 3 (skewed-high second, skewed-low third), fraction	0.30 (0.46)	0.24 (0.43)
Order 4 (skewed-high third, skewed-low second), fraction	0.17 (0.37)	0.26 (0.44)
Health context 1 (health gain in years), fraction	0.34 (0.48)	0.33 (0.47)
Health context 2 (health gain in months), fraction	0.33 (0.47)	0.34 (0.47)
Health context 3 (health gain in reduced hours of pain), fraction	0.33 (0.47)	0.34 (0.47)
Payoff to patients (0=hypothetical, 1=real)	0.49 (0.50)	0.50 (0.50)
Multiple switches in frame “high skew”, fraction	0.10 (0.30)	0.07 (0.26)
Multiple switches in frame “low skew”, fraction	0.08 (0.27)	0.12 (0.32)

Sample size, n=257. Mean and standard deviations in parentheses. All 257 individuals performed two frames: high skew and low skew.

Table 5. Distribution of Safe Choices by Health Context and Frames (in percentages)

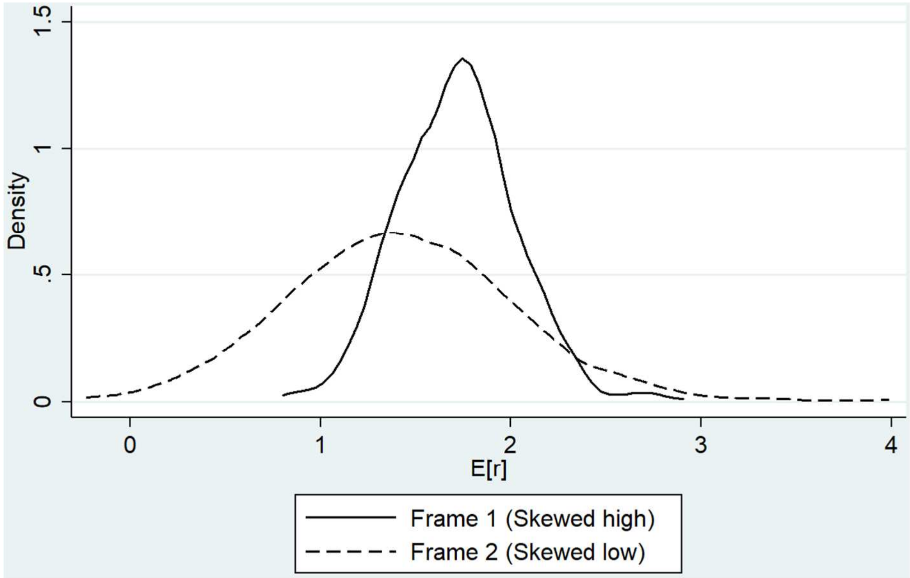
Number of safe choices	Health context 1 (years perfect health)		Health context 2 (months moderate ill-health)		Health context 3 (hours pain alleviated)	
	Frame “high skew”	Frame “low skew”	Frame “high skew”	Frame “low skew”	Frame “high skew”	Frame “low skew”
1	1.2	3.5	2.3	2.3	1.2	1.2
2	2.4	0.0	1.2	0.0	2.3	1.2
3	0.0	1.2	1.2	1.2	0.0	1.2
4	5.9	0.0	2.3	0.0	7.0	1.2
5	12.9	2.4	9.3	5.8	7.0	1.2
6	22.4	10.6	16.3	10.5	16.3	8.1
7	9.4	9.4	19.8	8.1	14.0	5.8
8	21.2	14.1	14.0	8.1	15.1	14.0
9	12.9	16.5	18.6	18.6	11.6	17.4
10	7.1	18.8	7.0	22.1	8.1	17.4
11	4.7	23.5	8.1	23.3	17.4	31.4
Total	100 (N=85)	100 (N=85)	100 (N=86)	100 (N=86)	100 (N=86)	100 (N=86)

Table 6. Interval Regression Model Estimation of Coefficient of Relative Risk Aversion

Variable	Skewed High (a)	Skewed Low (b)	Combined Frames (c)
Individual characteristics			
Age	-0.014	-0.037	-0.025
Gender (0=male, 1=female)	0.015	-0.287	-0.136
Health status (1=Excellent, 2=Very good, 3=Good, 4=Regular, 5=Poor)	-0.106	-0.091	-0.107
Labor activity last week, part time (less than 15 hours)	0.123	0.226	0.177
Labor activity last week, full time (more than 15 hours)	-0.240	-0.203	-0.249
Weekly disposable income (in Euros)	0.002	0.002	0.002
Health related major (0=no health related, 1=health related)	0.036	0.684***	0.308*
Year in college (1st to 7th year)	0.026	0.091	0.052
Grade of access to college (scale 0 to 1)	-0.218	-0.48-	-0.283
Number of people per room in household	0.400	0.795	0.603*
Education of father (0=none, 1=school, 2=college, 3=post-graduate)	0.045	0.089	0.056
Education of mother (0=none, 1=school, 2=college, 3=post-graduate)	0.040	-0.132	-0.038
Experimental design characteristics			
Order 2 (skewed high second, skewed low first)	-0.153	-0.360	-0.226
Order 3 (skewed high second, skewed low third)	0.154	0.149	0.148
Order 4 (skewed high third, skewed low second)	0.186	-0.535	-0.119
Health context 2 (health gain in months)	0.143	0.121	0.091
Health context 3 (health gain in reduced hours of pain)	0.315**	0.389	0.348**
Payoff to patients (0=hypothetical, 1=real)	-0.225*	-0.522**	-0.377***
Frame "low skew"			-0.384***
Multiple switches	0.237	-0.513	0.268
Constant	1.744***	2.01*	2.796***
Standard deviation of individual random effect			0.894***
Standard deviation of error term	0.944***	1.589***	0.910***

(a) Interval regression for skewed-high frame. Sample size is 254. Log-likelihood value is -528.37. (b) Interval regression for skewed-low frame. Sample size is 251. Log-likelihood value is -423.64. (c) Unbalanced panel data interval regression combining the skewed-high and skewed-low frames. A total of 255 individuals are included in the estimation. Total sample size is 505. Log likelihood value is -953.96. Wald test of null hypothesis that all coefficients are zero has a chi-square value of 51.58 (p-value<0.01). The fraction of the total variance error associated to random individual effect is 0.492 (standard error is 0.05). * Significant at 10%; ** significant at 5%; *** significant at 1%.

Figure 1. Kernel Distribution of Coefficient of Relative Risk Aversion by Frame



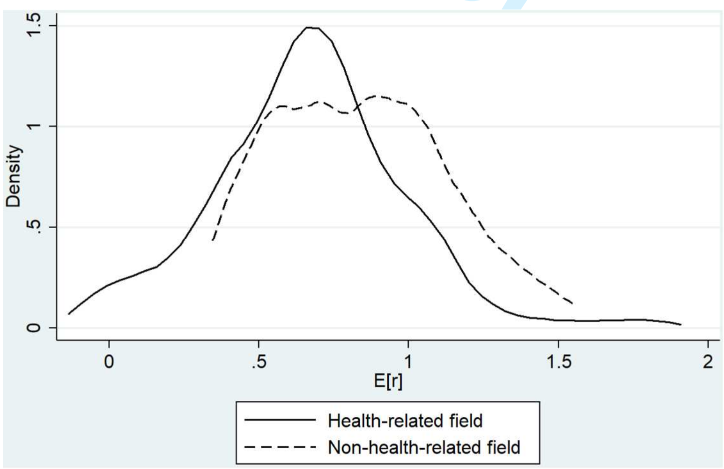
Probability density estimation based on an Epanechnikov kernel function. The figure represents the distribution of the coefficient of relative risk aversion, predicted by the Interval regression estimates, for the skewed-high frame (N=254) and the skewed-low frame (N=251).

Appendix A: Kernel distribution for H&L in the monetary domain and in the health domain

This appendix compares predicted coefficients of relative risk aversion based on H&L in the monetary domain and H&L in the health domain. Predicted coefficients were obtained from the interval regression estimations by i) type of students (health-related students versus non-health-related students), and ii) type of payoffs (real versus hypothetical payoffs).

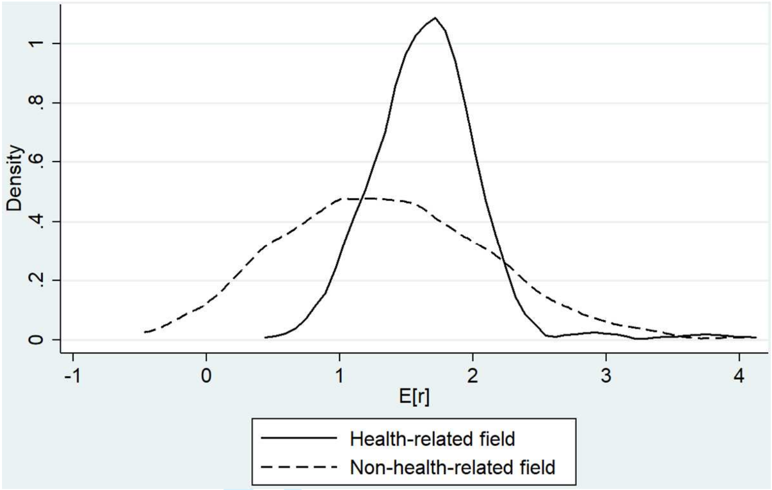
Figures A.1 and A.2 show the kernel distribution of the relative risk reduction coefficient, by type of student, for the monetary domain and the health domain, respectively. The results show that health-related students are less risk-averse than non-health related students in the monetary domain (r=0.665 Vs. r=0.832, mean coefficient of relative risk aversion of health-related Vs. non-health-related field, respectively), but more risk-averse in the health domain (r=1.654 Vs. r=1.336, mean coefficient of relative risk aversion of health-related Vs. non-health-related field, respectively).

Figure A.1. Kernel Distribution of Coefficient of Relative Risk Aversion: medical Vs. non-medical student in the monetary domain



Probability density estimation based on an Epanechnikov kernel function. The figure represents the distribution of the coefficient of relative risk aversion based on the H&L in the monetary domain. Coefficients are predicted by the Interval regression estimates, for students in a health-related field (N=178) and students in a non-health-related field (N=78).

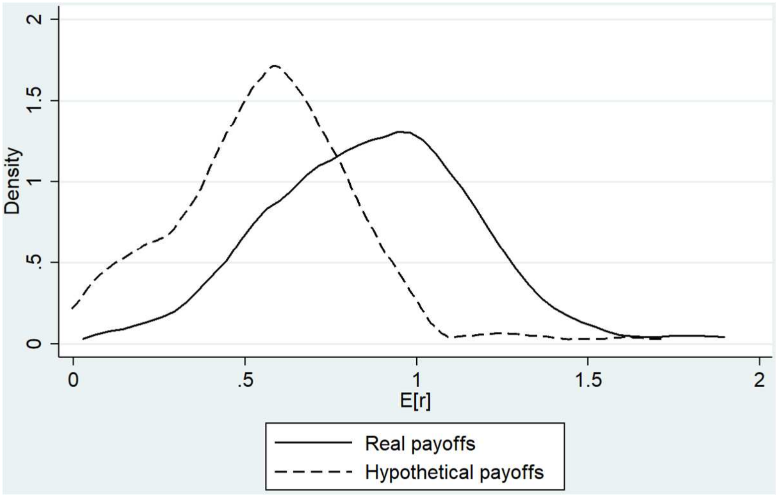
Figure A.2. Kernel Distribution of Coefficient of Relative Risk Aversion: medical vs. non-medical student in the health domain



Probability density estimation based on an Epanechnikov kernel function. The figure represents the distribution of the coefficient of relative risk aversion based on the H&L in the health domain. Coefficients are predicted by the Interval regression estimates, combining the skewed-high and skewed-low frames, for students in a health-related field (N=351) and students in a non-health-related field (N=154).

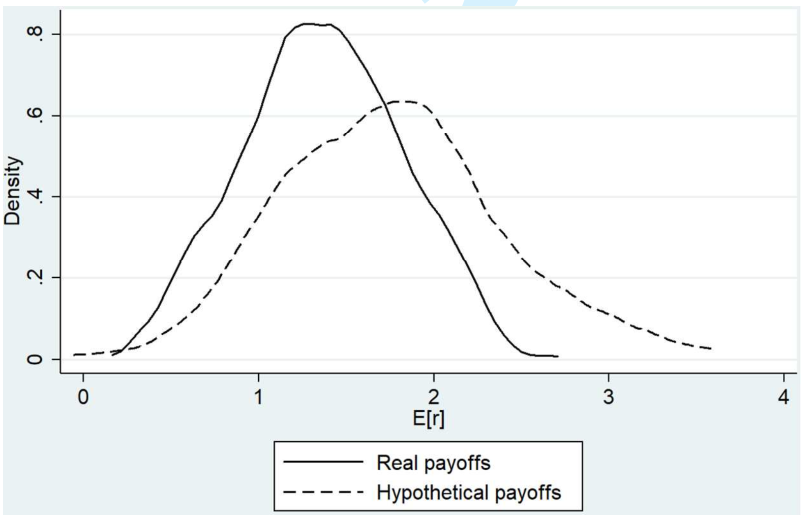
Figures A.3 and A.4 show the kernel distribution of coefficient of relative risk reduction, by type of payoffs, for the monetary domain and the health domain, respectively. The results show that subjects in the real payoffs are more risk-averse than those in the hypothetical payoffs when they are in the monetary domain (r=0.566 Vs. r=0.872, mean coefficient of relative risk aversion of real payoff group Vs. hypothetical payoff group, respectively), but more risk-averse in the health domain (r=1.756 Vs. r=1.365, mean coefficient of relative risk aversion of real payoff group Vs. hypothetical payoff group, respectively).

Figure A.3. Kernel Distribution of Coefficient of Relative Risk Aversion: hypothetical vs. real in the monetary domain



Probability density estimation based on an Epanechnikov kernel function. The figure represents the distribution of the coefficient of relative risk aversion, based on the H&L in the monetary domain. Coefficients are predicted by the Interval regression estimates, for subjects in the real payoffs group (N=128) and in the hypothetical payoff group (N=128).

Figure A.4. Kernel Distribution of Coefficient of Relative Risk Aversion: hypothetical Vs. real in the health domain



Probability density estimation based on an Epanechnikov kernel function. The figure represents the distribution of the coefficient of relative risk aversion, based on the H&L in the health domain. Coefficients are predicted by the Interval regression estimates, combining the skewed-high and skewed-low frames, for subjects in the real payoffs group (N=254) and in the hypothetical payoff group (N=251).

Appendix B: General Instructions for subjects in real rewards groups.[25]

[The general instructions were given to the participants before the start of the experiment. The original instructions were presented in Spanish; the following is a translated version. These instructions were identical for all groups with real rewards].

I N S T R U C TIONS

This is an experiment to study how individuals make decisions. We are only interested in what individuals make on average. Keep in mind that we do not expect any particular behavior from you. However, be aware that your decisions throughout the experiment may influence your earnings. You can ask questions at any time; please, raise your hand and wait to be helped. Any communication among you is prohibited and subject to immediate exclusion from the experiment.

- You are about to participate in a study funded by the Spanish Ministry of Science and Innovation. In this study we are analyzing some aspects of medical decision-making, which will give us a better understanding of the relationship between doctors and patients, and contribute to the design of appropriate health policies. Pay attention, please, to these instructions because your answers are very important. We thank you in advance for your cooperation.
- This experiment consists of three phases. In the first phase you will face a number of monetary decisions, while in the other two phases you will take health decisions. Before starting each phase, you will receive a set of specific instructions for that phase.
- At the completion of the three phases of the experiment, we will ask you to complete a short questionnaire and you will be announced the earnings that were generated with the decisions you made.
- By answering all our questions you will receive a payment of 5 Euros in cash at the end of the experiment. Also you can get additional money depending on the choices you make. At each phase we will explain what determines the payments and how they are realized.
- At the end of the experiment the computer will make three random selections (one for each of the three phases of the experiment) to determine payments based on the monetary and health decisions that you have made.

[25] Appendix C provides the general instructions and questionnaires to elicit risk aversion with money as outcome for those groups with real rewards. Appendix D provides the general instructions and questionnaires to elicit risk aversion with health as outcome for those groups with real rewards that were assigned to health context 1. Finally, Appendix E provides general instructions corresponding to the third part of the experiment whose data we do not analyze in this paper. Similar to Appendix C and D, Appendix E provides the questionnaires for groups with real rewards assigned to health context 1. The questionnaires for the groups with real rewards assigned to health contexts 2 and 3, as well as those corresponding to the hypothetical groups, are not included for space reasons but are available upon request. Each of the three different health contexts analyzed in the article are described in detail in section 3.3.

- As you can see, the gains from your financial decisions (Phase 1) will be for you, while the gains from your health decisions will be allocated to fund real medical treatments. In particular, the money will be transferred to Vicente Ferrer Foundation to finance the medical treatments offered by the network of health centers that the Foundation manages in Anantapur district, located in the state of Andhra Pradesh in India. Vicente Ferrer Foundation is an NGO committed to improving the development of the living conditions of the most disadvantaged communities in Andhra Pradesh.
- We have agreed with Vicente Ferrer Foundation that once completed the experiment, we will make a bank transfer for the total amount of the donation. Once the transfer has been made we will inform you by email.
- As everyone has the same opportunity to be paid for their decisions at the end of the experiment, we encourage you to choose the option you prefer at any phase, knowing that this option might be selected for the final payment.
- It is very important to know that this is not a test of skills since there is not a best option. The choice will depend on your personal preferences, which will not necessarily coincide with those of the rest of the participants.

Appendix C: multiple price list task with money as outcome for subjects in real rewards groups.

[These are the instructions for Part 1 of the experiment. The original instructions were presented in Spanish; the following is a translated version. These instructions were identical for all groups with real payoffs].

In this phase of the experiment you have to make 11 choices between two possible options (A and B). With Option A, you can receive 10 or 8 Euros. With Option B, you receive 19.25 or 0.5 Euros. As you can see, potential outcomes are fixed for the 11 choices, while the probability of gaining one amount or the other is different.

Next, we show the screenshot of the example that you will face at this phase.

Decision 1	**A:** 0% 10 euros , 100% 8 euros	**B:** 0% 19.25 euros, 100% 0.5 euros
Decision 2	**A:** 10% 10 euros, 90% 8 euros	**B:** 10% 19.25 euros, 90% 0.5 euros
Decision 3	**A:** 20% 10 euros, 80% 8 euros	**B:** 20% 19.25 euros, 80% 0.5 euros
Decision 4	**A:** 30% 10 euros, 70% 8 euros	**B:** 30% 19.25 euros, 70% 0.5 euros
Decision 5	**A:** 40% 10 euros, 60% 8 euros	**B:** 40% 19.25 euros, 60% 0.5 euros
Decision 6	**A:** 50% 10 euros, 50% 8 euros	**B:** 50% 19.25 euros, 50% 0.5 euros
Decision 7	**A:** 60% 10 euros, 40% 8 euros	**B:** 60% 19.25 euros, 40% 0.5 euros
Decision 8	**A:** 70% 10 euros, 30% 8 euros	**B:** 70% 19.25 euros, 30% 0.5 euros
Decision 9	**A:** 80% 10 euros, 20% 8 euros	**B:** 80% 19.25 euros, 20% 0.5 euros
Decision 10	**A:** 90% 10 euros, 10% 8 euros	**B:** 90% 19.25 euros, 10% 0.5 euros
Decision 11	**A:** 100% 10 euros, 0% 8 euros	**B:** 100% 19.25 euros, 0% 0.5 euros

- Please note that in Decision 1, option A implies 8 euros for certain, while option B ensures you 0.5 euros. At the same time, in Decision 11, option A guarantees 10 euros, while option B guarantees 19.25. Thus, as you can see, going down the table, option B becomes more attractive, as it increases the probability of winning the larger amount (19.25 Euros). Note that in option A the same thing happens.
- It is very important to know that this is not a test of skills since there is not a best option. The choice will depend on your personal preferences, which will not necessarily coincide with those of the rest of the participants.
- At the end of the experiment, the computer will randomly choose one of your 11 choices in part 1 to make your payment.

Payments in Part 1

Suppose, for instance, computer randomly selects Decision 5.

- *Option A*: 40% 10 Euros, 60% 8 Euros
- *Option B*: 40% 19.25 Euros, 60% 0.5 Euros

If you have chosen option A you can receive 10 or 8 euros. If you chose option B you can receive 19.25 or 0.5 euros.

Thus, afterwards, the computer will randomly select a number between 1 and 11. In both options A and B you have 40% of probability of receiving the larger amount of money in each of the options and 60% of receiving the lowest amount. Thus, if the computer randomly selects a number lower than 4 you will receive the maximum amount, while if the number is larger than 6 you will receive the minimum.

Suppose for instance that you have chosen Option A. If the computer selects a number between 1 and 4 (both included), you will receive 10 euros and if the number is between 5 and 10 you will receive 8 euros. Similarly, if you have chosen Option B, and the computer selects a number between 1 and 4 (both included), you will receive 19.25 euros and if the number is between 5 and 10 you will receive 0.5 euros.

Please select one of the options in each of the following 11 decisions.

End Phase I

Appendix D: Two multiple price list tasks with health as outcome for subjects in real rewards groups.

[These are the instructions for Part 2 of the experiment. The original instructions were presented in Spanish; the following is a translated version. These instructions were identical for all groups with real payoffs that were assigned to the health context 1 (health gains in years of life in perfect health), except for the order effects included between frames (skewed high or skewed low) and parts (tasks to elicit risk aversion with health as outcome versus task to elicit equity preferences). The results of the equity task will be analyzed in a different paper].

In phases II and III of the experiment we will ask you to imagine what you would do in a number of medical decisions. While it is true that in this experiment there are not real patients, we want you to think about patients as if they were real. In order to do this, the choices you make in these two phases will benefit real patients. In particular, the money earned with your decisions will fund treatments that Vicente Ferrer Foundation delivers to patients through its network of health centers in the district of Anantapur, in the state of Andhra Pradesh in India.

In this room there are 32 individuals (including you) facing exactly the same decisions. At the end of the experiment, the computer will randomly select one of the participants in order to compute the amount of the donation directed to Vicente Ferrer Foundation. The same will be done in Phase III. If you are chosen, thus, your decision will determine the health outcomes of your (hypothetical) patients, which in turn will imply a transfer of money to Vicente Ferrer Foundation.

Lastly, as perhaps it may seem that the amount of money you can give to Vicente Ferrer Foundation is not high, you should think that in India the cost of many basic treatments to reduce mortality is very low. For example with 5 Euros you can save the life of a patient with typhus.

Phase II

In this phase of the experiment we ask you to imagine what you would do in a number of medical decisions to treat a patient. As you will see, the situation we show you is very simple. We know that, in real life, medical decisions are much more complex. Try to respond what you believe you would do in each of the cases that are presented.

Let us start with an example of a Decision that you should take at this phase of the experiment.

Suppose you are a doctor you have to treat a middle-aged adult patient having a disease diagnosed with certainty. You have to choose between two possible treatments (A and B). In the case of treatment A, if it is successful, the patient will gain additional 16 years of life and if it does not work, 12 years. With treatment B, if it is successful, the patient will gain 39 additional years and if it goes wrong 4 years. **Importantly, the patient will enjoy good health the additional years of life**. However, the probability of going well or badly of each treatment is unknown. That depends on each patient. We present 11 different choices. In each, the probability that A and B treatments are successful is different. We ask you to tell us what treatment you choose in each of the 11 situations.

Next, we show a screenshot of one of the examples that you will face at this phase.

[TABLE CONTEXT a.1]

Decision 1	**A:** 0% 16 years of life , 100% 12 years	**B:** 0% 39 years of life, 100% 4 years
Decision 2	**A:** 10% 16 years, 90% 12 years	**B:** 10% 39 years, 90% 4 years
Decision 3	**A:** 20% 16 years, 80% 12 years	**B:** 20% 39 years, 80% 4 years
Decision 4	**A:** 30% 16 years, 70% 12 years	**B:** 30% 39 years, 70% 4 years
Decision 5	**A:** 40% 16 years, 60% 12 years	**B:** 40% 39 years, 60% 4 years
Decision 6	**A:** 50% 16 years, 50% 12 years	**B:** 50% 39 years, 50% 4 years
Decision 7	**A:** 60% 16 years, 40% 12 years	**B:** 60% 39 years, 40% 4 years
Decision 8	**A:** 70% 16 years, 30% 12 years	**B:** 70% 39 years, 30% 4 years
Decision 9	**A:** 80% 16 years, 20% 12 years	**B:** 80% 39 years, 20% 4 years
Decision 10	**A:** 90% 16 years, 10% 12 years	**B:** 90% 39 years, 10% 4 years
Decision 11	**A:** 100% 16 years, 0% 12 years	**B:** 100% 39 years, 0% 4 years

- Note that, although the mechanism is similar to the previous one, now what you win or lose is very different: years of life in perfect health instead of euros. Options A and B are different and therefore probably what you choose in each Decision may be different as well.
- Remember that at the end of the experiment, the computer will randomly choose one of the 32 participants to compute the payments of this phase of the experiment. Such payments will depend on the years of life gained by the patient of the selected participant. We proceed as before:
- The computer will choose the Decision that is played first.
- The computer randomly draws a number from 1 to 10 to determine the number of years of life that will gain your patient, given your selected treatment.

- Depending on the years in perfect health the patient gained, the donation to Vicente Ferrer Foundation will be higher or lower. In particular, to make payments at this phase, we assume that an additional <u>year of life in perfect health</u> amounts to 4 Euros.
  - If the patient gains 39 years, Vicente Ferrer Foundation will receive 156 Euros.
  - If the patient gains 16 years Vicente Ferrer Foundation will receive 64 Euros.
  - If the patient gains 12 years Vicente Ferrer Foundation will receive 48 Euros.
  - If the patient gains 4 years Vicente Ferrer Foundation will receive 16 Euros.

In what follows we ask you to tell us which treatment you would choose in each of the following 11 decisions.

[INSERT TABLE CONTEXT a.1]

The frame we present now also refers to **the number of years of life in perfect health** that a patient can gain. However, the numbers have changed and, therefore, your preferred option for each Decision may change too. We ask you to tell us which treatment you would choose in each of these 11 situations that we show below.

If any of them is chosen by the computer to make the payment, the computation is performed as in the previous example. In this case, the Vicente Ferrer Foundation will receive 76 Euros, 64 Euros, 48 Euros or 16 Euros, if your patient gains 19, 16, 12 or 4 years of life in perfect health respectively.

[TABLE CONTEXT a.2]

Decision 1	**A:** 0% 16 years of life , 100% 12 years	**B:** 0% 19 years of life, 100% 4 years
Decision 2	**A:** 10% 16 years, 90% 12 years	**B:** 10% 19 years, 90% 4 years
Decision 3	**A:** 20% 16 years, 80% 12 years	**B:** 20% 19 years, 80% 4 years
Decision 4	**A:** 30% 16 years, 70% 12 years	**B:** 30% 19 years, 70% 4 years
Decision 5	**A:** 40% 16 years, 60% 12 years	**B:** 40% 19 years, 60% 4 years
Decision 6	**A:** 50% 16 years, 50% 12 years	**B:** 50% 19 years, 50% 4 years

Decision 7	**A:** 60% 16 years, 40% 12 years	**B:** 60% 19 years, 40% 4 years
Decision 8	**A:** 70% 16 years, 30% 12 years	**B:** 70% 19 years, 30% 4 years
Decision 9	**A:** 80% 16 years, 20% 12 years	**B:** 80% 19 years, 20% 4 years
Decision 10	**A:** 90% 16 years, 10% 12 years	**B:** 90% 19 years, 10% 4 years
Decision 11	**A:** 100% 16 years, 0% 12 years	**B:** 100% 19 years, 0% 4 years

End Phase II

Appendix E: Equity task for subjects in real rewards groups

[These are the instructions for Part 3 of the experiment. The original instructions were presented in Spanish; the following is a translated version. These instructions were identical for all groups with real payoffs that were assigned to health context 1 (health gains in years of life in perfect health), except for the order effects included between frames (skewed high or skewed low) and parts (tasks to elicit risk aversion with health as outcome versus tasks to elicit equity preferences). The results of this equity task will be analyzed in a different paper.]

Phase III

In phase III of the experiment we will ask you to imagine what you would decide in a number of medical situations involving the treatment of a **group of patients**. As you will see, the situation we show you in each case is very simple. We know that in real life medical decisions are much more complex. Try to respond according to what you believe you would do in each of the cases that are presented.

Suppose you are responsible for deciding which treatment should be provided to a population of middle-aged adult patients that have been diagnosed with certainty. Your choice will be the treatment protocol all physicians in your area will follow. There are two types of patients, type A and type B, both suffering from the same illness. You have to choose between two possible treatments (A and B). Treatment A is better for type A patients, while treatment B is more appropriate for type B patients. That is:

Increase in life expectancy		Treatment	
		A	B
Type of patient	A	16 years of life	4 years of life
	B	12 years of life	39 years of life

- If a type A patient receives treatment A, then his life expectancy increases by 16 years.
- If a type B patient receives treatment A, then his life expectancy increases by 12 years.
- If a type A patient receives treatment B, then his life expectancy increases by 4 years.
- If a type B patient receives treatment B, then his life expectancy increases by 39 years.

The patient will enjoy good health during the additional years of life.

Notice that in this scenario:
1. There is no way to know a patient's type.
2. Once you provide one treatment, you cannot provide the other.

We present 11 different choices with different population compositions. For simplicity, we assume there are 100 patients. We ask you to tell us what treatment the health system should choose to treat these patients in each of the 11 situations.

Next, we show the screenshot of the decisions that you will face at this phase.

	Patient composition	Treatment A	Treatment B
Decision 1	100 Type A	All patients gain 16 years	All patients gain 4 years
Decision 2	90 Type A 10 Type B	90 patients A gain 16 years 10 patients B gain 12 years	90 patients A gain 4 years 10 patients B gain 39 years
Decision 3	80 Type A 20 Type B	80 patients A gain 16 years 20 patients B gain 12 years	80 patients A gain 4 years 20 patients B gain 39 years
Decision 4	70 Type A 30 Type B	70 patients A gain 16 years 30 patients B gain 12 years	70 patients A gain 4 years 30 patients B gain 39 years
Decision 5	60 Type A 40 Type B	60 patients A gain 16 years 40 patients B gain 12 years	60 patients A gain 4 years 40 patients B gain 39 years
Decision 6	50 Type A 50 Type B	50 patients A gain 16 years 50 patients B gain 12 years	50 patients A gain 4 years 50 patients B gain 39 years
Decision 7	40 Type A 60 Type B	40 patients A gain 16 years 60 patients B gain 12 years	40 patients A gain 4 years 60 patients B gain 39 years
Decision 8	30 Type A 70 Type B	30 patients A gain 16 years 70 patients B gain 12 years	30 patients A gain 4 years 70 patients B gain 39 years
Decision 9	20 Type A 80 Type B	20 patients A gain 16 years 80 patients B gain 12 years	20 patients A gain 4 years 80 patients B gain 39 years
Decision 10	10 Type A 90 Type B	10 patients A gain 16 years 90 patients B gain 12 years	10 patients A gain 4 years 90 patients B gain 39 years
Decision 11	100 Type B	All patients gain 12 years	All patients gain 39 years

At the end of the experiment, the computer will randomly choose one of the 32 participants to compute the payments of this phase of the experiment. If you are selected, the computer will randomly choose the decision that will determine the consequences of your choice (treatment A or B). As before, the choices you make in this phase will benefit real patients. In particular, the money earned with your decisions will fund treatments in 10 rural health centers that Vicente Ferrer Foundation manages in the district of Anantapur. We assign a number to each health care center in order to assign the amount of money to be transferred to each of them. In particular, the donation that will be received by each of the health care centers will be based on the health gains obtained by every ten patients.

Suppose for instance that the computer randomly selects the following decision:

Decision 2	90 Type A 10 Type B	90 patients A gain 16 years 10 patients B gain 12 years	90 patients A gain 4 years 10 patients B gain 39 years

In this case:

- If you have chosen treatment A, 90 patients gain 16 years of life and 10 patients 12 years. Thus, 9 health centers (those numbered 1 to 9) will receive 16 euros and health center number 10 will receive 12 euros.
- If, on the contrary, you choose treatment B, then 90 patients gain 4 years of life and 10 patients gain 39. In this case health centers numbered 1 to 9 will receive 4 euros and health center number 10 will receive 39 euros.

Once the experiment is over, we will send to the Vicente Ferrer Foundation the money and the instructions corresponding to the allocation of funds for each of the health care centers.

In what follows we ask you to tell us which treatment you would choose in each of the following 11 decisions.

	Patients' composition	Treatment A	Treatment B
Decision 1	100 Type A	All patients gain 16 years	All patients gain 4 years
Decision 2	90 Type A 10 Type B	90 patients A gain 16 years 10 patients B gain 12 years	90 patients A gain 4 years 10 patients B gain 39 years
Decision 3	80 Type A 20 Type B	80 patients A gain 16 years 20 patients B gain 12 years	80 patients A gain 4 years 20 patients B gain 39 years
Decision 4	70 Type A 30 Type B	70 patients A gain 16 years 30 patients B gain 12 years	70 patients A gain 4 years 30 patients B gain 39 years
Decision 5	60 Type A 40 Type B	60 patients A gain 16 years 40 patients B gain 12 years	60 patients A gain 4 years 40 patients B gain 39 years
Decision 6	50 Type A 50 Type B	50 patients A gain 16 years 50 patients B gain 12 years	50 patients A gain 4 years 50 patients B gain 39 years
Decision 7	40 Type A 60 Type B	40 patients A gain 16 years 60 patients B gain 12 years	40 patients A gain 4 years 60 patients B gain 39 years
Decision 8	30 Type A 70 Type B	30 patients A gain 16 years 70 patients B gain 12 years	30 patients A gain 4 years 70 patients B gain 39 years

Decision 9	20 Type A 80 Type B	20 patients A gain 16 years 80 patients B gain 12 years	20 patients A gain 4 years 80 patients B gain 39 years
Decision 10	10 Type A 90 Type B	10 patients A gain 16 years 90 patients B gain 12 years	10 patients A gain 4 years 90 patients B gain 39 years
Decision 11	100 Type B	All patients gain 12 years	All patients gain 39 years

End Phase III