**Working papers series**

**WP ECON 06.16**

# *The influence of the Ratio Bias phenomenon on the elicitation of Standard Gamble utilities*

Jose Luis Pinto Prades (U. Pablo de Olavide)

Jorge E. Martinez Perez (U. of Murcia)

Jose Maria Abellan Perpiñan (U. of Murcia)

JEL Classification numbers: I10.
Keywords: framing effect, risk format, standard gamble, health state, dual-process theories.

UNIVERSIDAD PABLO DE OLAVIDE SEVILLA

**Department of Economics**

# The influence of the Ratio Bias phenomenon on the elicitation of Standard Gamble utilities

Jose Luis Pinto Prades*
University Pablo de Olavide
Sevilla, Spain

Jorge E. Martinez Perez
University of Murcia
Murcia, Spain

Jose Maria Abellan Perpiñan
University of Murcia
Murcia, Spain

Abstract

This paper tests whether logically equivalent risk formats can lead to different health state utilities elicited by means of the standard gamble (SG) method. We compare SG utilities elicited when probabilities are framed in terms of frequencies with respect to 100 people in the population (i.e., X out of 100) with SG utilities elicited for frequencies with respect to 1,000 people in the population (i.e., Y out of 1,000). We found that utilities were significant higher when success and failure probabilities were framed as frequencies type "Y out of 1,000" rather than as frequencies type "X out of 100". This framing effect, known as Ratio Bias, may have important consequences in resource allocation decisions.

KEYWORDS: framing effect, risk format, standard gamble, health state, dual-process theories.

ACKNOWLEDGEMENTS: we would like to thank John Baron, Han Bleichrodt, Seymour Epstein, Robin Hogarth, Graham Loomes, Richard Smith and Kimihiko Yamagishi for their helpful comments. The usual disclaimer applies.

*Corresponding author: Department of Economics, University Pablo de Olavide, Sevilla, Spain

# 1. Introduction

'Framing' refers to the wording and/or other means to present logically equivalent information. A 'framing effect' (Tversky and Kahneman, 1981) arises when alternative framing of this logically equivalent information produces different decisions. Evidence to date suggests that this framing effect is a common phenomenon affecting both hypothetical and real decisions made by patients and physicians (see Edwards et al., 2001, for a review).

One aspect of information that is often presented is risk – the probability of an event, such as a cure or side-effect resulting from treatment. In the case of risk information, empirical evidence suggests several types of framing effect. First, preference reversals may occur when equivalent risk information is presented in a negative or a positive frame (i.e. as a gain or a loss) (Eraker and Sox, 1981; O'Connor et al., 1985; O'Connor, 1989; Banks et al., 1995; Redelmeier et al., 1993; Llewellyn-Thomas et al., 1995; Gurm and Litaker, 2000; Amstrong et al., 2002). Second, quite different decisions over alternative treatments can emerge when the same information is presented as relative risk, absolute risk, or the number-needed-to treat (Forrow et al., 1992; Malenka et al., 1993; Bucher and Weinbacher, 1994; Hux et al., 1994; Sarfati et al., 1998; McGettigan et al., 1999). Third, numerically equivalent risk formats (e.g. frequency versus percentage) can lead to inconsistent preferences (Slovic et al., 2000, Schapira et al., 2004). Some research has indicated that frequency formats are more easily understood than probability formats (e.g., Bowling and Ebrahim, 2001). The use of frequency formats is supported by those authors (Gigerenzer and Hoffrage, 1995; Hoffrage and Gigerenzer, 1998; Hoffrage et al., 2000) that have suggested that, for diagnostic inference, such as the determination of pre- and post-test probabilities, frequency formats are easier for people to use than percentage formats. However, some research has shown (Denes-Raj and Epstein, 1994) that some people are also subject to biases using frequency formats. The "ratio-bias phenomenon" (RBP) is a paradigmatic case.

The RBP suggests that when having to evaluate the likelihood of an event, subjects have a tendency to focus on the numerator, disregarding the denominator. This bias has not received much attention in medical decision making despite its potential relevance. For example, some (e.g. Barrat et al, 2005) describe the consequences of screening mammography in terms of number of cancers diagnosed in a group of 1000 women, while others (UHMS, 2004) describe the outcomes in terms of events in a group of 10,000 women, and yet others as the number of cancers diagnosed in a population of 100,000 women (NBCC, 2002). What the RBP suggests is that the potential benefit of mammography screening will look different in these three cases even if the likelihood of events is exactly the same.

The objective of this paper is to investigate whether, and how, the RBP may influence utility derivation. Specifically, to test if utilities elicited through the standard gamble (SG) method are susceptible to the RBP. This is important to determine, as the SG method is quite frequently used to elicit von-Neumann Morgenstern utilities for individual decision making and has been used to estimate population utilities in instruments like the Health Utility Index (Torrance et al., 1995).

The paper is structured as follows. In section 2 we describe the RBP and offer a psychological explanation for it. In section 3 we hypothesize that SG utilities elicited when probabilities are framed in terms of frequencies with respect to 1,000 people in the population will be higher than SG utilities elicited for frequencies with respect to 100 people in the population. This prediction is based on the joint effect of, at least, four motivations for the presence of the RBP phenomenon: small number effect, numerosity, saliency and the affect heuristic. Section 4 describes the methods and results of two experiments designed to test the potential influence of the RBP in SG. The paper closes with a discussion of the implications of the findings from this experiment.

## 2. Definition and explanation of the Ratio-Bias Phenomenon

In another study Denes-Raj and Epstein (1994) showed that, when offered a chance to win $1 by drawing a red jelly bean from an tray, people often chose to draw from a bowl containing a greater absolute number, but a smaller proportion, of read beans (9 out of 100) than from a bowl with fewer red beans but a better chance of winning (1 out of 10). In the case of the win trials, 61% of subjects chose the large tray offering a smaller chance of winning (9% vs 10%) and in the case of the lose trials 39% of subjects chose the smaller tray offering a higher chance of a loss. When participants were asked to justify their choice, they admitted that this choice went contrary to what a rational individual should do, but they felt they had a better chance when there were more red beans.

Other research has shown that this bias is not a mere curiosity of laboratory experiments with students. Slovic et al (2000) showed that 40% of clinicians refused to discharge a mental patient when violence risk was communicated as “20 out of every 100 patients similar to Mr. Jones are estimated to commit an act of violence”, but only 20% refused to discharge the patient when risk was explained as “2 out of every 10 patients similar to Mr. Jones are estimated to commit an act of violence”. Yamagishi (1997a) found a clear inconsistency in risk judgements provided by lay people using frequency formats. In his experiment participants rated a disease that kills 1,286 people out of every 10,000 as more dangerous than one that kills 24.14 out of every 100. The RBP has also been produced using vignettes [Denes-Raj, V., Epstein, S., and Cole, J. (1994)].

Numerosity refers to the tendency to judge quantity or probability on the basis of the number of units into which a stimulus is divided without fully considering variables like the size of the units (Pelham et al., 1994). Pacini and Epstein (1999) use this heuristic in order to explain the RBP. They claim that the experiential system encodes and better comprehends absolute numbers (numerosity) than ratios because single numbers are more concrete than relations between numbers. However, this heuristic does not explain why people focus on the number in the numerator and not in the denominator. One reason provided by Denes-Raj and Epstein (1994) is that the numerator is the object of motivational concern. For example, the red beans are of motivational concern since we want to select them (if the outcome is a gain) or avoid them (if the outcome is a loss). The final question is, why people focus on *one aspect* of the numerator (e.g., number of red jelly beans, number of criminals that commit violent actions, number of people who die) and not in the other aspect (e.g. number of *white* jelly beans, number of criminals that *do not* commit violent actions, number of people who *do not* die)? One argument that Kirkpatrick and Epstein (1992) provide is that red jelly beans are more salient because there are less of them, standing out as figure against ground. Another argument provided in Pacini and Epstein (1999) is related to the fact that instructions are stated in terms of drawing a red jelly bean.

The small numbers effect asserts that the experiential system comprehends smaller numbers better than larger numbers. In this respect, a 1 in 10 probability conveys the idea of a low probability better than 10 in 100, because subjects find it easier to visualize 10 than 100 subjects. The smaller tray gives a better idea than the larger tray that a 10% chance is a low probability. For this reason, if the lottery is positive (win if red, nothing if white) people choose the larger tray and the opposite if the lottery is negative. Some evidence that probabilities expressed with low numbers are easier to interpret is provided in Pelham et al. (1994) where subjects expressed their preferences about participating in a lottery where *all participants* are given one ticket or in another lottery where *all participants* are given 10 tickets. People tend to choose the lottery where they are given 10 tickets if they are told that there are 1 million participants but they show indifference between both lotteries when they are told that there are only 2 participants. Apparently, with lower numbers they better realize that the chances are the same in both lotteries.

Slovic et al. (2000) explain the results by Kirkpatrik and Epstein (1992) and Denes-Raj and Epstein (1994) as a manifestation of a mental strategy of “imaging the numerator” (the red beans) and “neglecting the denominator” (number of beans in the tray). According to Slovic et al., images of winning beans convey positive affect that motivates the choice of the bowl with the greater absolute number of red beans. This mechanism was coined by Slovic et al. (2002) as the *affect heuristic*. Applying this to their paper they say that 20 mental patients out of 100 conducting violent acts may evoke more images of harmful attacks than 2 out of 10. The role of affect in explaining the RBP also fits within the framework provided by dual process theories since the experiential system influences, and is influenced by, affect. Affect determines what is attended to and what is reinforced by the experiential system (Epstein, 2003).

Yamagishi (1997a) suggests a combination of two cognitive mechanisms as an explanation to his result (why a risk of 12.86% is perceived as higher than a risk of 24.84%): *anchoring and adjustment* (Tversky and Kahneman, 1974) and *base-rate neglect* (Kahneman and Tversky, 1973). Yamagishi argues that people use the numerator as an anchor to do subsequent judgment and, simultaneously, tend to reject the base rate. As 1,286 is higher than 24.14 then a risk of 1,286 out of 10,000 is perceived as more dangerous than a risk of 24.14 out of 100, even though the former probability (12.86%) is lower than the latter one (24.84%). The problem is to explain why people anchor and adjust from one part of the numerator (the number of bad events) and do not pay enough attention to the denominator. The arguments provided above (numerosity, saliency, affect) can explain this.

We would like to add a further reason to explain the Slovic et al (2000) and Yamagishi (1997a) results, which is that they only provided information about the number of people that could commit violence acts or the number of people who could die. For example, Slovic et al (2000) did not explain to their subjects that 8 out 10 (or 80 out of 100) would *not* commit any act of violence. This way of presenting information may have lead subjects to focus on this part of the numerator. Again, this argument is just a description of a fact but the relevant question is why this happens; why people apparently do not place enough attention to information (number of people that do not die or that do not commit violence acts) that, although it is not explicitly provided, it is implicit in the data. We propose two explanations for this effect. One is the motivational concern already mentioned; the number of people who die or the number of people who commit violence acts is the main motivational concern. The fact that the researchers only provide explicit information about the number of *positive* events (people who do something) and not on the number of *negative* events (people who do not do something) exacerbates the motivational concern. The second also relates to the fact that the experiential system takes a leading role in these kinds of questions. As the experiential system is characterized by instinctual thinking we believe that it is easier for the experiential system to focus on occurrences (things that happen, like having an accident) than on non-occurrences (things that do not happen, like not having an accident). If this is true, even if we provide information in a balanced way, even if we mention explicitly the number of good and bad events, subjects will focus on the number of bad events.

In summary then, there are several reasons that can explain the RBP:

1.-Numerosity: people understand numbers more easily than ratios.

Saliency: people concentrate on the number of events that are less frequent.

3.-Small numbers. People have a better idea of the likelihood of an event expressed in small numbers.

4.-Biased Instructions: asking people to think in the consequence of a certain kind of event (e.g. drawing a red ball).

5.-Neglect of the base: tendency to forget about the denominator.

6.-Affect heuristic. People respond by imaging the part of the numerator that is more salient. This effect reinforces saliency but needs the former.

7.-Partial information: they only give explicit information about one part of the numerator (e.g. number of people that can commit violence acts and not those that will not commit those acts).

As we have seen, all these reasons have an underlying explanation: all of them are congruent with the idea that the experiential system has a leading role in this kind of tasks. We know proceed to apply this theory to the case of SG utilities.

## 3. Hypotheses and tests

*Hypotheses*

We can use the above explanations of the RBP in order to generate our hypotheses about the potential influence of this bias on SG utilities.

If we elicit SG utilities asking people the risk of death that they would accept in order to improve their health, then all seven reasons provided above will influence the response and utilities will be higher if frequencies are expressed using 1000 as the denominator (rather than 100). The reason is that this type of question will make people focus on the number of deaths and since this number is higher with 1000 as denominator, subjects will accept lower risks and they will produce higher utilities than where the denominator is 100. However, since we are interested in eliciting unbiased utilities, we try to avoid some of the reasons that give rise to the RBP. For this reason:

1. We try to frame the decision in an otherwise neutral way, such that people are not led to concentrate on only one outcome (probability of death). We then elicit utilities using a sequence of choices. In this way, people do not have to focus on one single component of the numerator. That is, we want to find out $N_{100}$ and $N_{1000}$ such that the lotteries [D, $N_{100}$; (FH, RL)] and [D, $N_{1000}$; (FH, RL)] have the same utility as the sure outcome (Q, RL). If we ask people which value of $N_{100}$ or $N_{1000}$ makes both prospects equivalent we are making people focus only on one single number, namely, $N_{100}$ ($N_{1000}$). However, if we ask people to choose between (Q, RL) and [D, $N_{100}$; (FH, RL)] or [D, $N_{1000}$; (FH, RL)] and we keep moving both probabilities (of D and of FH) so that we finally reach indifference, we will reduce the chances that people just concentrate on $N_{100}$ ($N_{1000}$).
2. We provide information about both potential outcomes. That is, we provide information about $N_{100}$ and (1- $N_{100}$). The sequence of choices that we have described will present subjects with both pieces of information.

3. We provide subjects with visual aids to help them think in terms of the total denominator so it is as clear as possible that the absolute number of good or bad events is related to the total number of potential events.

However, in spite of making every effort to provide information in the most unbiased way possible, the reasons behind the RBP cannot be totally eliminated. More specifically, it is hard to imagine how to avoid the "saliency" effect and the "small-number" effect (and consequently the affect heuristic). If the event with a lower frequency is more salient, there will always be a "saliency" effect except in the case of events with a 50% chance. If the "small-number" effect is an intrinsic part of human thinking, there will always be a "small-number" effect. This leads us to propose two specific hypotheses:

1. *Existence of the RBP:* due to the effect of numerosity, small-numbers, saliency and the affect heuristic, SG utilities will be higher when the likelihood is framed in terms of X out 1000 people in the population than when it is framed in terms of X out of 100 people. This is because: A. Smaller numbers convey better the idea of small risk. B. Numerosity leads people to concentrate on absolute numbers and not ratios. C. Saliency leads people to concentrate on events that are less frequent (number of deaths). D. Affect will work more strongly when the absolute number of deaths is higher (although the probability is the same). For this reason, we predict that people will accept a lower risk if frequencies are provided in terms of N out of 1000 than if they are provided in terms of N out of 100, leading to higher SG utilities. The hypothesis about the existence of the RBP will be tested comparing the SG utilities elicited from one group (henceforth, $U_{SG100}$) and SG utilities elicited from the other group ($U_{SG1,000}$). Our prediction is that $U_{SG1,000}>U_{SG1,00}$.
2. *Intensity of the RBP:* according to the "saliency" hypothesis, the influence of the RBP in SG utilities will be larger for milder health states than for more severe ones. This is because the saliency of the event "death" will be higher in the milder health states since there will be less bad events (deaths) for milder health states. There are two ways of interpreting the intensity of the RBP. One is to assume that the higher the effect the higher the <u>difference</u> $U_{SG1,000}-U_{SG1,00}$. We would then expect this difference to be greater for milder health states. However, this test may run into problems since for the mild health state

the difference between the utility of the health state and the utility of full health can be very small. For example, if $U_{100}$=0.99 then the RBP would have "no room" to show up. So another possible way of testing the intensity of the RBP is through the ratio:

$$\frac{1-U_{1000}}{1-U_{100}} \qquad (1)$$

For example if $U_{100}$=0.9 and $U_{1000}$=0.99 we would consider that the RBP has a higher effect than if $U_{100}$=0.5 and $U_{1000}$=0.59. In any case, in order to study the relevance of the RBP for economic evaluation of health care technologies, (1) is the key ratio since it is the ratio of the benefits of medical treatments.

*Tests*

We split a convenience sample into two groups. We asked participants in one group to set probabilities that yield indifference between the gamble [D, $N_{100}$; (FH, RL)] and in the other for the gamble [D, $N_{1000}$; (FH, RL)] and sure outcome (Q, RL). Overall, four SG questions were presented to each subject, one for each of the health states selected. In one group 100 was used as the denominator and in the other group 1,000 was used.

## 4. The experiment

- Sample

The participants were 200 economics students at the University of Murcia (Spain). They were paid 6 Euro for their participation. Responses were collected by face-to-face interview, with the questionnaire pilot-tested prior to the actual experiment.

- Health states

We used the EQ-5D health states 22111, 11222, 22222, and 23222. These states are described in Table 1. Throughout the experiment, the health states were labeled health state X, W, Z and Y respectively.

[Insert table 1 about here]

Given the ordinal structure of the component dimensions in the EuroQol descriptive system, some states are logically ordered with respect to others. With the states used here, five such comparisons are possible. It would be expected that 22222 should be given a higher utility than 23222 because it is better on at least one dimension and no worse on any of the other dimensions. In the same way, this ordinal consistency would be expected for comparisons of 22111 vs 22222, 22111 vs 23222, 11222 vs 22222, and 11222 vs 23222. Only for comparison between 22111 and 11222 there is no *a priori* expectation of this kind.

- Design

The experiment was run on a computer, which facilitated the use of visual aids and the choice-based SG procedure to elicit utilities. In addition, the pilot sessions showed that people found computer assisted personal interviews a user-friendly procedure.

To avoid anchoring biases we split the total sample into two groups of 100 subjects each. One group answered four SG questions (one per health state) in which probabilities were framed in terms of frequencies with respect to 1,000 people in the population, while the other group answered the same questions for probabilities framed as frequencies with respect to 100 people in the population. To avoid order effects, the computer randomnly varied the order in which the different SG questions were asked. To minimize response errors, subjects had to confirm the elicited indifference value after each question. As a preliminary task, participants were asked to rate the health states on a visual analogue scale (VAS), with 100 (best imaginable health state) and 0 (worst imaginable health state) as endpoints. The principal objective of the VAS was to familiarize participants with the health state descriptions and also to have a test that both groups were similar.

Recruitment of participants took place one week before the actual experiment started. At recruitment, participants were handed a practice question. Participants were asked to answer this practice question at home. This procedure was intended to familiarize participants with the SG questions. Prior to the start of the experiment, subjects were asked to explain their answer to the practice question. When we were not convinced that a subject understood the task, we explained it again until we were convinced that he understood the task.

The formulation of SG questions was the same, regardless of the frequency format used. For example, in the case of health state X, the wording of the question (translated from the Spanish) was as follows:

> *Suppose that you are experiencing health state X. If you do not receive treatment you will remain in X for the rest of your life. However, you can receive a medical treatment (ALFA treatment), that if successful, will result in return to normal health. Nevertheless, ALFA treatment can also fail and in this case you will die. We are going to show you different probabilities of success and failure and you will tell us if you think you would choose treatment ALFA or no in each caset.*

The SG procedure was then applied to elicit utilities. Frequencies were displayed using human figures, as frequency formats illustrated with human figures have been shown to be easy to interpret and convey a meaningful message (Schapira et al., 2001). Figures 1 and 2 illustrate the way indifferences were obtained.

[Insert figures 1 and 2 about here; visual aid]

Suppose that for the probabilities displayed in figure 1 (5% D, 95% FH), the individual prefers treatment ALFA to no treatment. Next the computer would display a new choice (figure 2), where probabilities of success and failure would be 10 in 100 and 90 in 100 respectively. Suppose that for these probabilities the individual prefers the sure outcome rather than treatment ALFA. Then the computer would present the next choice (10%, D; 90%FH) vs no treatment. This iterative procedure would continue until indifference is reached.

- Data analysis and statistical methods

As normality was rejected by the Kolmogorov-Smirnov test with Lilliefors correction ($p < 0.0001$), we used nonparametric procedures in order to test for the significance of differences between SG utilities elicited for frequencies with respect to 100 people in the population and SG utilities elicited with respect to 1,000 people in the population.

We also tested whether the two distributions of SG utilities differed with respect to median, using the Wilcoxon Mann-Whitney test. This is one of the most powerful of the non-parametric tests for comparing two independent samples (Hollander and Wolfe, 1999).

- Results

Table 2 shows means, medians and standard deviations (SD) corresponding to the VAS for each sample.

[Insert table 2 about here]

Both samples gave similar values and no significant differences were detected ($\alpha$=0.05) between medians. It was then concluded that both groups have similar preferences for health states. Also in both samples the rankings were internally consistent in the sense that state Y receives a lower value than state Z, and both Y and Z receive lower values than states X and W. This shows that subjects understood the basic task of valuing health states [A-HA!].

Table 3 shows the main results. Utilities elicited from the group in which frequencies were described as “Y out of 1,000” are consistently higher than utilities elicited from the group in which frequencies were described as “X out of 100”. This tendency is robust for all health states. This was our prior expectation and confirms our hypothesis about the existence of the RBP. The second prediction, namely, a stronger effect for milder health states depends on the test that we apply. Differences between $U_{1000}$ and $U_{100}$ are almost constant and in this sense the strength of the effect is constant. However, the ratio of equation (1) is lower for milder health states. This implies that the RBP will distort more the utilities of mild health states than the utilities of severe health states, at least if they are used for economic evaluation of health care technologies.

[Insert table 3 about here]

## 5. Further evidence

Once we had seen the influence of the RBP on the SG utilities we wanted to see if this phenomenon was also present in other contexts and in other populations more relevant for public policy. We then conducted a second study that had some differences with respect to the first study:

1. Subjects were members of the general population.
2. The tasks to be performed by subjects were the same that had been previously used in a study addressed to estimate the value of a statistical life (VSL) from road accidents in Britain.
3. The tasks were a modified SG question characterized by the existence of risk in both parts of the question.

### - The sample

We used two sub-samples of the Spanish general population (n=180 each). They were chosen using a quota sample method. No statistically significant differences were found between both sub-samples in socio-demographic characteristics. Responses were collected by face-to-face interviews that were held at subjects home.

### - Health states

We used the same health states used by Carthy et al. (1999) in their study of the VSL. They are described in Table 4. One difference with our first study is that they did not involved chronic illnesses.

[Insert table 4 about here]

- Tasks

Subjects had to conduct two modified SG questions. Both tasks were quite similar. First, they were told that they had to suppose that they had been injured in a road accident. If untreated the injury would result in death. They were informed that there were two alternative treatments available. One treatment, if successful would result in the hospitalization and prognosis associated with injury W, but if unsuccessful would result in immediate unconsciousness followed shortly by death with probability 1 in 1000. The second treatment, if successful would result in a return to normal health (NH) within 3-4 days, but if unsuccessful would result in immediate unconsciousness followed shortly by death with probability $\Psi$ (> 1 in 1000). So the first task can be represented as estimating [$\Psi$, Death; (1- $\Psi$), NH] such that it produces the same utility level than [0.001, Death; 0.999, W]. The second task was similar to this one. In second case, respondents were asked to assume that without treatment they would experience the hospitalization and prognosis associated with injury X. They were also informed that there were two alternative treatments available. One treatment, if successful would result in the hospitalization and prognosis associated with injury W, but if unsuccessful would result the hospitalization and prognosis associated with injury X with probability 1 in 100. The second treatment, if successful would result in a return to normal health (NH) within 3-4 days, but if unsuccessful would result in the hospitalization and prognosis associated with injury X (> 1 in 100). So the second task can be represented as estimating [$\rho$, X; (1- $\rho$ ), W] such that it produces the same utility level than [0.01, X; 0.99, NH].

We asked both questions to both groups. However, while question 1 was framed exactly the same in both groups, question 2 was framed using different denominators. In group A, the risk was framed in terms of 1 out 100. In group B, the risk was framed in terms of 10 out 1000. So question 2 in group A was framed as the value of $\rho$, that made indifferent gamble [$\rho$, X; (1- $\rho$ ), W] to gamble [1 in 100, X; 99 in 100, NH]. In group B was framed as the value of ρ, that made indifferent gamble [ρ, X; (1- ρ), W] to gamble [10 in 1000, X; 990 in 1000, NH]. Visual aids similar to those used in experiment 1 were also used.

- Hypothesis

According to the RBP as people concentrate in the risk of the bad outcome of the numerator, they would accept a lower risk ρ in group B (denominator 1000) than in group A (denominator 100).

- Results

Table 5 shows means and medians corresponding to both groups. As suggested by the RBP, ρ was lower in group B. So while subjects in group A accepted a risk of 37 out of 100, subjects in group B only accepted a risk of 176 out of 1000.

[Insert table 5 about here]

## 6. Discussion.

This paper provides new evidence on how 'irrelevant' changes in the way we represent health risks can lead to inconsistent preferences. We find that utilities were significant higher when risk information was framed using 1,000 instead of 100 as denominator.  It shows how superficially different frequency frames (e.g. X out of 100 vs Y out of 1,000) can distort SG measurements.

How relevant is this bias for practical decision making? Well, let's assume we use 1000 as the denominator when considering the benefit of curing somebody in health state Y, which is more than 5 times the benefit of curing somebody in X  [(1-0.481)/(1-0.902)=5.27]. However, this relative benefit almost halves [(1-0.393)/(1-0.790)=2.89] if we use 100 as the denominator. Since there is no reason to prefer frequencies in 100 or in 1000, there is no reason to necessarily choose one framing over the other. Of course, if one wishes to deliberately increase the apparent relative benefit of Y respect to X, one can use 1000 as the denominator for Y and 100 as the denominator for X. This increases the relative benefit of Y to more than six times the benefit of X.  For individual decision making, it remains to be seen how this bias can influence medical decisions since there has not been any study, to our knowledge, that has investigated this issue.

Having seen the potential implications of this bias, what can we do in order to avoid it? This depends on the perspective taken about the origins of the RBP. The RBP observed in SG values can be explained by the dual-process theory, such that SG is handled more by the experiential than the rational system. This system works using heuristics that, although can do a good job in some decisions, may lead to irrationalities, like choosing a dominated lottery. In the case of preference elicitation for health states, the experiential system leads people to focus on just one component of the likelihood of one event. This is logically irrational. However, we do not imply that the rational system is totally absent in the elicitation task. For example, the rational system may work trying to avoid irrational responses, like accepting a higher risk of death for a health state that is less severe than another. This leads to utilities that are internally consistent, as ours. For this reason, we can see at the same time consistent (e.g. higher utilities for milder health states) but biased responses. This is similar to the idea behind the principle of coherent arbitrariness (Ariely et al., 2003), that apparently rational decisions can be based on arbitrary "anchors".

If we accept that the above account of the origin of the RBP bias is correct, how can we deal with it? One approach, in the spirit of Bleichrodt et al (2001) is to correct the "irrational" SG utilities. Some authors have shown that the SG method is affected by loss aversion and probability weighting (Bleichrodt, 2001, 2002). Since we have a good understanding of these biases there are quantitative corrections to these biases (Bleichrodt et al., 2001; van Osch et al., 2004; Bleichrodt et al., 2005). For example, it has been suggested that utilities should be estimated under Prospect Theory and not under Expected Utility. In this way, the bias coming from probability transformation can be avoided. This approach assumes that the rational system wants to adhere to expected utility and then corrects the experiential system in order to choose according to a rational rule. While in the case of some biases affecting SG (loss aversion, probability transformation) this could be the case, we cannot see how this correction can be applied to the RBP, since EU has nothing to do with the choice of the denominator.

Slovic et al (2000) suggest a pragmatic strategy. They suggest using several formats in order to communicate risk. Translating this to our case, we need to provide information using several denominators. Of course, this complicates the task of eliciting utilities. In order to avoid this complication we may try to confirm the hypothesis that people understand small numbers more than large numbers. If this is the case we would undertake the elicitation task using small numbers, as far as possible. For example, instead of asking subjects if they would accept a risk of death of X out of 100, it would be better to use X out of 10. If people say that they would accept 1 out of 10 but not 2 out of 10, we know that the indifference point is between 10% and 20%. We can the use larger numbers (X in 100, whereas 10<X<20 in our example) if we need a more accurate estimate. This approach reflects the Bleichrodt et al (2001) approach in the sense that it assumes there is a denominator that is normatively better than others; the smaller one. It could also be that this is a useful short term solution, and in the longer term it might be interesting explore ways to "educate" the experiential system so that these biases can be reduced or eliminated. For this, it would be quite helpful to categorize subjects according to how they handle conflicts between the experiential and the rational system. It has been shown [Epstein et al (1996), Pacini and Epstein (1999), Frederick (2005), Amstel et al (paper under review)] that some subjects are more prone to intuitive thinking and biased judgements than others. If this was the case, it would have important implications for individual and social decision making. In individual decision making, those subjects that are more susceptible to biases could be made aware of this problem in order that avoid taking biased decisions. In social decision making, it should be discussed if preferences of these subjects should or should not be taken into account in order to design social policy.

In conclusion, researchers working in utility measurement should be aware of the existence of this bias since it may have relevant consequences for individual and social decision making. We also believe that the RBP can be quite useful in understanding better the origin of other biases since it may help to study the way that the rational and the experiential system work when subjects respond to preference elicitation tasks.

References

Amsel, E., Close, J., Sadler, E., & Klaczynski, P. (Under Review). Awareness and irrationality: College students' awareness of their irrational judgments on gambling tasks. *Journal of Gambling Studies.*

Ariely, D., Loewenstein, G. & Prelec, D. (2003). Coherent Arbitrariness: Stable Demand Curves without Stable Preferences. *Quarterly Journal of Economics*,118 ,73-105

Armstrong, K., Schwartz, J.S., Fitzgerald, G., Putt, M. & Ubel, P. (2002). Effect of framing as gain versus loss on understanding and hypothetical treatment choices: survival and mortality curves. *Medical Decision Making,* 22, 76-83.

Banks, S.M., Salovey, P., Greener, S., Rothman, A.J., Moyer, A., Beauvais, J. & Epel, E. (1995). The effects of message framing on mammography utilization. *Health Psychology* ,14, 178-184.

Barratt, A., Howard, K.,Irwig, L., Salkeld, G. &, Houssami, N. (2005). Model of outcomes of screening mammography: information to support informed choices. *British Medical Journal*, 330, 936-938.

Bleichrodt, H. (2001). Probability weighting in choice under risk: an empirical test. *Journal of Risk and Uncertainty,* 23, 185-198.

Bleichrodt, H. (2002). A New Explanation for the Difference Between Standard Gamble and Time Trade-Off Utilities. *Health Economics,* 11, 447-456.

Bleichrodt, H., Pinto, J.L. & Wakker, P. (2001). Making Descriptive Use of Prospect Theory to Improve the Prescriptive applications of Expected Utility. *Management Science*, 47, 1498-1514

Bleichrodt, H., Abellan-Perpiñan, J.M., Pinto, J.L. & Mendez-Martinez, I. (2005). Resolving inconsistencies in utility measurement under risk: tests of generalizations of expected utility. Working Paper UPF 718.

Bostic, R., Hernstein, R.J. & Luce, R.D. (1990). The effect on the preference reversal phenomenon of using choice indifferences. *Journal of Economic Behavior and Organization* , 3, 193-212.

Bowling, A. &, Ebrahim, S. (2001). Measuring patients' preferences for treatment and perceptions of risk. *Quality in Health Care* ,10 ,2-8.

Bucher, H.C., Weinbacher, M. & Gyr, K. (1994). Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration, *British Medical Journal,* 309, 761-764.

Carthy, T., Chilton, S., Covey, J., Hopkins, L., Jones-Lee, M., Loomes, G., Pigdeon, N. & Spencer, A. (1999). On the contingent valuation of safety and the safety of contingent valuation: part 2- The CV/SG "chained" approach. *Journal of Risk and Uncertainty* ,17, 187-213.

Denes-Raj, V. & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judment. *Journal of Personality and Social Psychology* , 66, 819-829.

Denes-Raj, V., Epstein, S., & Cole, J. (1994). The generality of the ratio-bias phenomenon. *Personality and Social Psychology Bulletin*, 10, 1083-1092.

Edwards A.G.K, Elwyn, G., Covey, J., Mathews, E., Pill, R. (2001). Presenting risk information: a review of the effects of 'framing' and other manipulations on patient outcomes. *Journal of Health Community* , 6, 61-82.

Epstein, S. (1983). The unconscious, the preconscious and the self-concept. In Suls,J. & Greenwald,S. (Eds.), *Psychological perspectives on the self*, Vol. 2 (pp 219-247). Hillsdale, NJ: Erlbaum.

Epstein, S. (2003). "Cognitive-experiential self-theory." In Millon,T & Lerner, M. J. (Eds.), *Comprehensive handbook of psychology*, Vol. 5 (pp 159-184): Personality and Social Psychology. Hoboken, NJ: John Wiley & Sons, Inc.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology*, 71, 390-405.

Eraker, S.A., Sox, H.C. (1981). Assessment of patients' preferences for therapeutic outcomes. *Medical Decision Making,* 1, 29-39.

Forrow, L., Taylor, W.C. &, Arnold, R.M. (1992). Absolutely relative: how research results are summarized can affect treatment decisions. The *American Journal of Medicine* , 92,121-124.

Frederick, S. (2005). Cognitive reflection and decision making, *Journal of Economic Perspectives*, 19 (4), 25-42.

Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychology Review* ,102, 4-704.

Gurm, H.S. & Litaker, D.G.(2000). Framing procedural risks to patients: is 99% safe the same as a risk of 1 in 100?. *Academic Medicine* , 75, 840-842.

Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine* ; 73, 538-540.

Hoffrage, U., Lindsey, S., Hertwig, R. & Gigerenzer, G. (2000). Communicating statistical information, *Science* 290, 2261-2262.

Hogarth, R. (2005). "Deciding analytically or trusting your intuition? The advantages and disadvantages of analytic and intuitive thought." In Haberstroh, S. & Betsch, T. (Eds.), *The Routines of Decision Making (pp 67-82).* Mahwah, NJ: Lawrence Erlbaum Associates.

Hollander, M. & Wolfe, D. (1999). Nonparametric Statistical Methods. New York: John Wiley & Sons, Inc.

Hux, J.E., Levinton, C.M. & Naylor, C.D. (1994). Prescribing propensity: influence of life-expectancy gains and drug costs. *Journal of General and Internal Medicine* , 9, 195-201.

Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychology Review*, 80, 237-251.

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

Kirkpatrick, L.A. & Epstein, S. Cognitive-experiential self-theory and subjective probability: further evidence for two conceptual systems. *Journal of Personality and Social Psychology* , 63, 534–544.

Llewellyn-Thomas, H.A., McGreal, J. & Thiel, E.C. (1995). Cancer patients' decisión making and trial-entry preferences: the effects of "framing" information about short-term toxicity and long-term survival. *Medical Decision Making* , 15, 4-12.

Luce, R.D. (2000). Utility of gains and losses: Measurement-theoretical and experimental approaches. New Jersey: Lawrence Erlbaum Associates, Inc.

Malenka D.J., Baron, J.A., Johansen, S., Wahrenberger, J.W. & Ross, J.M.(1993). The framing effect of relative and absolute risk. *Journal of General and Internal Medicine* , 8, 543-548.

McGettigan, P., Sly, K., O'Connell, D., Hill, S. & Henry, D. (1999). The effects of information framing on the practices of physicians. *Journal of General and Internal Medicine* , 14, 633-642.

National Breast Cancer Coalition (NBCC). (2002). The Mammography Screening Controversy: Questions and Answers. Retrieved March 10, 2006 from http://www.natlbcc.org/bin/index.asp?strid=498&depid=9&btnid=2.

O'Connor, A.(1989), Effects of framing and level of probability on patients' preferences for cancer chemotherapy. *Journal of Clinical Epidemiology* , 42,119-126.

O’Connor, A.M., Boyd, N.F., Tritchler, D.L., Kriukov, Y., Sutherland, H. & Till, J.E. (1985). Eliciting preferences for alternative cancer drug treatments. The influence of framing, medium, and rater variables. *Medical Decision Making* , 5, 453-63.

Pacini, R., & Epstein, S. (1999a). The interaction of three facets of concrete thinking in a game of chance. *Thinking and Reasoning*, 5, 303–325.

Pacini, R., & Epstein, S. (1999b). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, 76, 972-987.

Pelham, B. W., Sumarta, T. T., & Myaskovsky, L. (1994). The easy path from many to much: The numerosity heuristic. *Cognitive Psychology*, 26, 103-133.

Redelmeier, D.A., Rozin, P., & Kahneman, D. (1993). Understanding patients’ decisions: cognitive and emotional perspectives. *Journal of the American Medical Association* 270, 72-6.

Sarfati, D., Howden-Chapman, P., & Woodward, S.C. (1998). Does the frame affect the picture? A study into how attitudes to screening for cancer are affected by the way benefit are expressed. *Journal of Medical Screening*, 5, 137-140.

Schapira M.M., Nattinger, A.B., & McHorney, C.A. (2001). Frequency or probability? A qualitative study of risk communication formats used in health care. *Medical Decision Making,* 21, 459-467.

Schapira, M.M., Davids, S.L., McAuliffe, T.L., & Nattinger, A.B. (2004). Agreement between scales in the measurement of breast cancer risk perceptions. *Risk Analysis*, 24, 665-673.

Sloman, S. A. (1996). The Empirical Case for Two Systems of Reasoning. *Psychological Bulletin,* 119(1), 3-22

Slovic, P., Finucane, M.L., Peters, E., & MacGregor, D.G. (2002). The affect heuristic. In Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). *Heuristic and biases: The Psychology of Intuitive Judgement* 2002 (pp. 397-420). New York: Cambridge University Press.

Slovic, P., Monahan, J., & MacGregor, D.G. (2000). Violence risk assessment and risk communication: the effects of using actual cases, providing instructions and employing probability versus frequency formats. *Law and Human Behavior,* 24, 271-296.

Torrance, G.W., Furlong, W., Feeny, D., & Boyle, M. (1995). Multi-attribute preference functions:Health Utilities Index. Pharmacoeconomics, 7(6), 503-520.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: heuristics and biases. *Science,* 185, 453-458.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the pshychology of choice. *Science,* 211, 453-458.

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

Van Osch, S., Wakker, P.P., van den Hout, W.B., & Stiggelbout, A.M. (2004). Correcting biases in standard gamble and time tradeoff utilities. *Medical Decision Making*, 24, 511-517.

University of Michigan Health System (UMHS) (2004). Adult preventive health care: cancer screening. Ann Arbor (MI): University of Michigan Health System. Retrieved March 10, 2006 from http://cme.med.umich.edu/pdf/guideline/cancerscreening04.pdf

Yamagishi, K. (1997a). When a 12.86% mortality is more dangerous than 24.14%: implications for risk communication. *Applied Cognitive Psychology,* 11, 495-506.

Yamagishi, K. (1997b). Upward versus downward anchoring in frequency judgements of social facts. *Japanese Psychological Research,* 2, 124-129.

Sloman, S.A. (1996). The Empirical Case for Two Systems of Reasoning. *Psychological Bulletin*, 119(1), 3-22

Figures and tables

Table 1. Health states used in the study

| Health state X | Health state W |
|---|---|
| Some problems walking about<br>Some problems with performing self care activities (e.g. eating, washing or dressing)<br>No problems with performing usual activities (e.g. work, study, housework, family or leisure activities)<br>No pain or discomfort<br>Not anxious or depressed | No problems walking about<br>No problems with performing self care activities (e.g. eating, washing or dressing)<br>Some problems with performing usual activities (e.g. work, study, housework, family or leisure activities)<br>Moderate pain or discomfort<br>Moderately anxious or depressed |
| **Health state Z** | **Health state Y** |
| Some problems walking about<br>Some problems with performing self care activities (e.g. eating, washing or dressing)<br>Some problems with performing usual activities (e.g. work, study, housework, family or leisure activities)<br>Moderate pain or discomfort<br>Moderately anxious or depressed | Some problems walking about<br>Unable to wash or dress self<br>Some problems with performing usual activities (e.g. work, study, housework, family or leisure activities)<br>Moderate pain or discomfort<br>Moderately anxious or depressed |

Figure 1. Example of choice-based procedure (1)

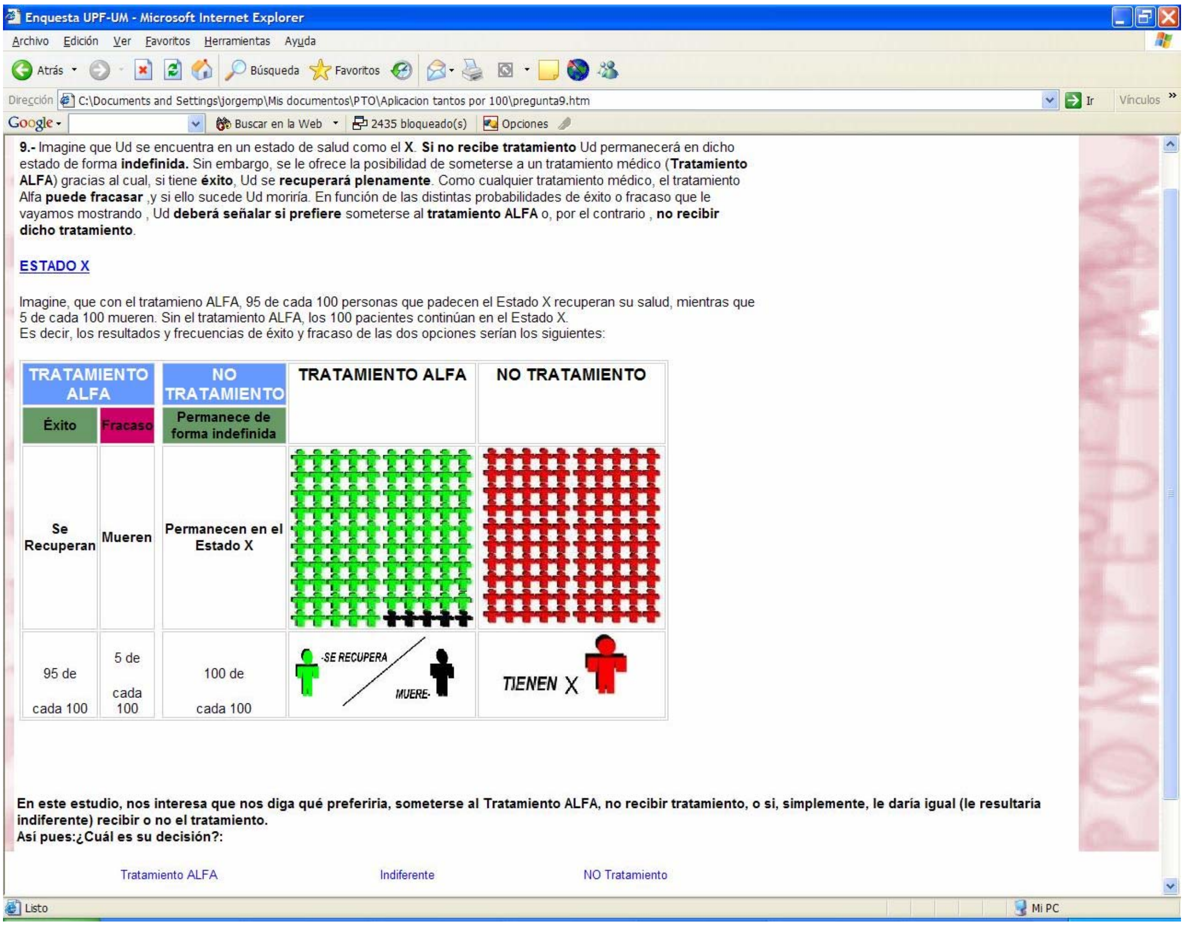

Figure 2. Example of choice-based procedure (2)

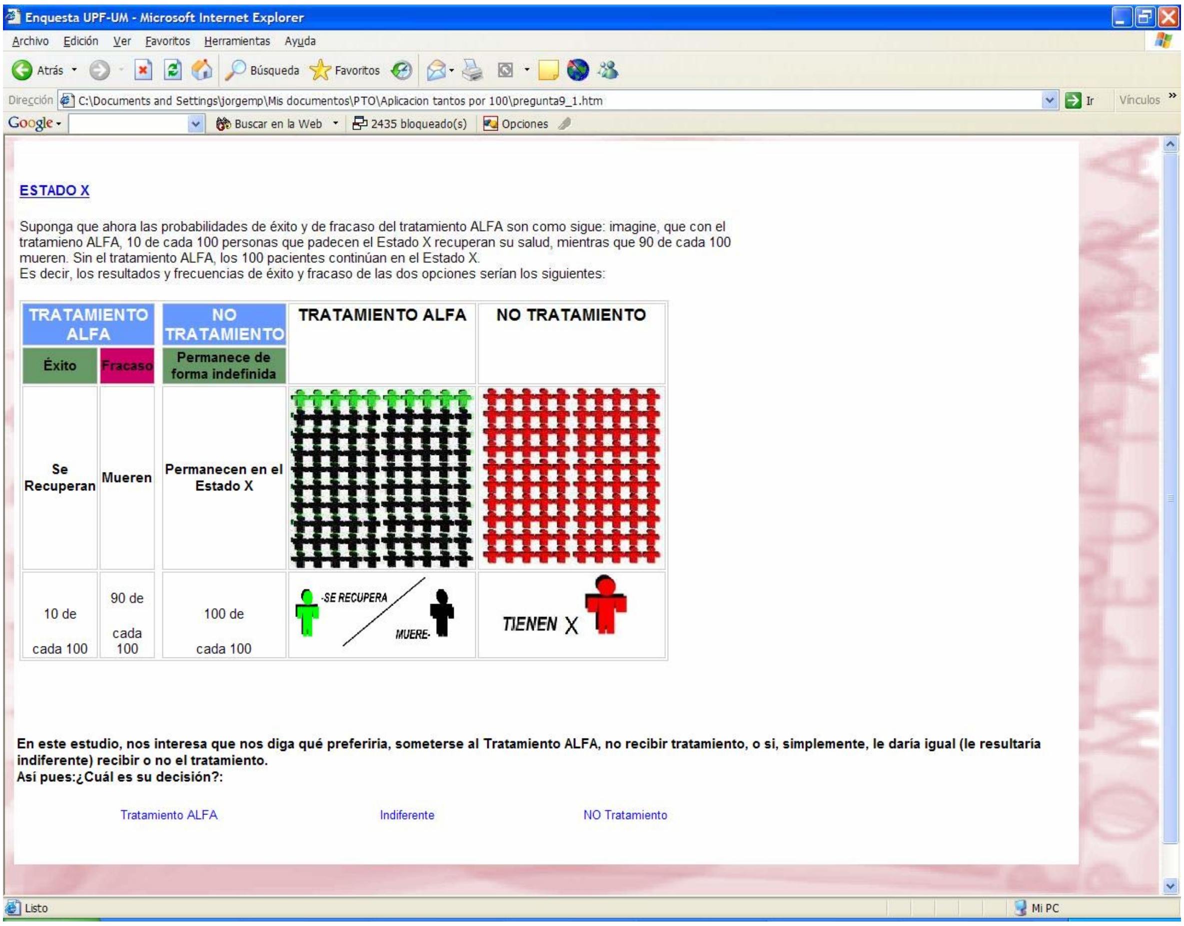

Table 2. Means, medians, standard deviation (SD) elicited by the VAS (subsamples, N=100).

| | Health state X | | Health state W | | Health state Z | | Health state Y | |
|---|---|---|---|---|---|---|---|---|
| | Out of 100 | Out of 1,000 | Out of 100 | Out of 1,000 | Out of 100 | Out of 1,000 | Out of 100 | Out of 1,000 |
| Mean<br>Median<br>(SD) | 57.32<br>60.00<br>(1.66) | 61.10<br>60.00<br>(1.57) | 56.23<br>60.00<br>(1.49) | 53.75<br>50.00<br>(1.47) | 30.56<br>30.00<br>(0.95) | 30.15<br>30.00<br>(1.17) | 16.62<br>15.00<br>(0.82) | 16.51<br>12.50<br>(0.94) |

**Note: differences between both groups were not statistically significant at the 5% level.**

Table 3. Utilities elicited by the SG (each sub-samples, N=100). Standard deviation (SD)

| | Health state X** | | Health state W** | | Health state Z* | | Health state Y* | |
|---|---|---|---|---|---|---|---|---|
| | Out of 100 | Out of 1,000 | Out of 100 | Out of 1,000 | Out of 100 | Out of 1,000 | Out of 100 | Out of 1,000 |
| Mean<br>Median<br>(SD) | 0.790<br>0.850<br>(0.185) | 0.902<br>0.945<br>(0.105) | 0.796<br>0.800<br>(0.161) | 0.880<br>0.908<br>(0.120) | 0.575<br>0.600<br>(0.271) | 0.673<br>0.700<br>(0.232) | 0.393<br>0.375<br>(0.272) | 0.481<br>0.500<br>(0.313) |
| $U_{1000}$-$U_{100}$ (from means) | 0.112 | | 0.084 | | 0.098 | | 0.088 | |
| $\frac{1-U_{1000}}{1-U_{100}}$ | 0.47 | | 0.59 | | 0.77 | | 0.86 | |

**** Significant different at α=0.01.**

- Significant different at α=0.05.

Table 4. Injury description cards.

| Injury X | Injury W |
|---|---|
| In hospital.<br>▪ 2 weeks<br>▪ slight to moderate pain.<br><br>After hospital<br>▪ some pain/discomfort, gradually reducing.<br>▪ some restrictions to work and leisure activities, steadily improving.<br>▪ after 18 months, return to normal health with no permanent disability | In hospital<br>▪ 2-3 days<br>▪ slight to moderate pain.<br><br>After hospital<br>▪ some pain/discomfort for several weeks.<br>▪ some restrictions to work and/or leisure activities for several weeks/months<br>▪ after 3-4 months, return to normal health with no permanent disability |

Table 5. Means, medians, T-test and Mann-U by groups.

| | Group A | | Group B | | | |
|---|---|---|---|---|---|---|
| Risk | Mean | Median | Mean | Median | T-test | Mann-U |
| ψ | 15.2 out of 1000 | 5 out of 1000 | 14.8 out of 1000 | 5 out of 1000 | 0.911 | 0.120 |
| ρ | 37.1 out of 100 | 33.5 out of 100 | 176 out of 1000 | 95 out of 1000 | 0.000 | 0.000 |