



EKONOMI  
HÖGSKOLAN  
Lunds universitet

Department of economics

Bachelor Thesis

Spring 2007

# Gambling with Health

A Method to Test Theories of Decision under Risk in the Domain  
of Health.

Supervisor: Krister Hjalte

Author: Jonatan Tops

# Abstract

In order to know how to efficiently allocate resources in health care we need a measure of its outcome, health. There is an ongoing debate on how the values of different health states are best measured. The dominating approach is to elicit utilities from decisions made by the public, in an experimental setting. In order to elicit valid utilities a theory of how decisions are made is needed. Traditionally the expected utility theory has been used but recently the prospect theory has been shown much interest. In this thesis a method is developed to test which one of these two theories is most suitable to use in the domain of health. The general idea behind the method is that each respondent constructs her own scale for rating health states. The method is also applied in a small scale experiment. It is concluded that the method seems applicable. Furthermore, the experiment indicates that probability weighting is plausible while the other features of the prospect theory are not clearly supported. The thesis ends with a brief discussion about how the method could be used in future surveys.

# Index

Abstract.....	2
1. Introduction.....	4
1.1 Issue .....	4
1.2 Purpose.....	6
1.3 Outline of the Thesis.....	6
1.4 Important Concepts.....	7
1.4.1 Health.....	7
1.4.2 Probability.....	7
1.4.3 Prospect.....	8
1.4.4 Risk and Uncertainty.....	8
2. The Theories .....	9
2.1 Notation.....	9
2.2 Expected Utility Theory.....	10
2.3 Standard Gamble.....	11
2.4 The Prospect Theory.....	12
2.5 Cardinality of Health.....	15
3. A Method of Investigation.....	18
3.1 The experiment .....	18
3.2 Learning from Observations .....	19
3.2.1 Knowing ones preferences.....	19
3.2.2 Answering Truthfully.....	20
3.2.3 Understanding the questions .....	20
3.3 Attributes.....	21
3.4 The Test Procedure .....	22
3.5 Validity .....	23
4. Experiment.....	26
4.1 Sample.....	26
4.2 The Health State Profiles .....	27
4.3 Validity of the Experiment.....	27
4.4 The Gambles .....	28
4.5 Results from the Experiment .....	29
4.6 Conclusions from the Experiment .....	32
5. Conclusions.....	35
5.1 Discussion.....	36
6. Appendix.....	37
7. References.....	39

# 1. Introduction

## 1.1 Issue

Finding the value of different health states is a field of research that is of great importance for medical decision-making at both an individual and an aggregate level. For example, if you want to allocate the health care resources where the marginal benefit is the largest you need to know how to measure the benefit; how much utility is produced by spending another dollar on a certain therapy. For the last decade it has been much debate on how the elicitation of these utilities should be carried out. Much of the attention has been directed towards methods in which people are asked to make choices that are then interpreted by the researcher in order to establish what value the respondent attaches to the different objects of choice. The interpretation of choices requires the researcher to use a theory of how people choose; he must assume something about how the choices are made in general to be able to conclude how the objects of choice are valued in relation to each other. The theory used to interpret the choices must then be an accurate description of how people actually choose otherwise the utilities elicited will be flawed. The reason that the respondent made a specific choice will not be that he valued the objects of choice in the way that the researcher concludes but that he made the decision in a way that the researcher did not foresee.

The standard gamble is one experimental setting used to elicit utility of health states. It is considered to have an advantage compared to other methods because it incorporates elements of risk. Since risk is generally involved in medical decision-making this is seen as a virtue (Torrance in Oliver, 2003). When one wants to interpret the choices made in

such a setting one is required to use a theory that describes decisions under risk in a realistic way. The standard gamble is, however, rooted in the Expected Utility Theory (EUT) which is now widely seen as an inadequate descriptive theory of decisions under risk (see e.g. Starmer, 2000, Machina, 1984, Camerer, 1998). The most prevalent non expected utility theory of decision under risk is the Prospect Theory (PT) (Kahneman& Tversky 1979; Tversky&Kahneman 1992). The PT was designed to explain the descriptive flaws of EUT and it is widely used to explain behavior under risk that is inconsistent with EUT (see e.g. Camerer, 1998). It has already been shown that the internal consistency<sup>1</sup> of standard gamble can be improved using PT (Bleichardt et. al. 1999, 2001, Oliver, 2003, Doctor et al. 2004). That gives some support to the idea that PT is a good description of decision under risk regarding health but it is not a decisive argument since it is used ex post. To understand the shortcomings of such an argument, imagine that you want to explain why water is transparent. You could argue that water is transparent since it is liquid, the theory (all liquids are transparent) does explain the phenomenon but that does not imply that the theory is correct. The theory has to be tested itself, for example by studying other liquids.

Furthermore, Kahneman&Tversky stated that “Although the present paper has been concerned mainly with monetary outcomes, the theory is readily applicable to choices involving other attributes, e.g., quality of life...” (Kahneman&Tversky 1979, p.288). However, this statement shall not be taken as a sound scientific argument but rather as an opinion.

Left to decide, before PT is accepted as the theory of decision under risk regarding health, is thus whether people follow the basic principles of PT also when the monetary outcomes are replaced by health states.

---

<sup>1</sup> Put simply, a method is internally consistent if it always gives the same ranking of the objects of choice. If you use the theory to interpret the choices to predict choices, then the more accurate the predictions are the better the internal consistency. A precise measure is given in Oliver (2003).

## 1.2 Purpose

The aim of this thesis is to develop a method to test if PT is a better descriptive theory than EUT also in the domain of health. The final part of the process to develop the method is to apply it in a small scale study.

The general question is: Is there a method to test the descriptive capacities of PT and EUT in the domain of health? Regarding the small scale experiment the answers to two questions are sought. Can EUT explain the decisions? And if not, is PT a better description of decision under risk in the domain of health than EUT?

## 1.3 Outline of the Thesis

The outline of this thesis will be as follows: First, important concepts will be defined and the theories that are later tested will be presented. Then a method intended to test the descriptive properties of PT and EUT is developed by identifying some fundamental requirements for an experimental method to work and then formulate a test procedure that satisfies these requirements. After this the results of a small scale empirical study, where the method is utilized, are described and conclusions about the descriptive capacity of EUT and PT are drawn. Lastly, general conclusions about the method are drawn and possible applications are discussed.

## 1.4 Important Concepts

There are some concepts that frequently are used in this thesis that need to be clarified.

### 1.4.1 Health

The most fundamental concept in this thesis is health. Does health exist in an objective sense so that we can observe it or is it just a matter of opinion? Economists have seen health as a good and as an investment (Grossman in Folland et al. 2004). In this thesis health will be dealt with as a composite good. We define health as  $H = \{a, b, c \dots\}$ , where  $a, b, c \dots$  are different attributes that together constitute the concept of health. This is a view that is implicitly assumed by the EuroQoL group by using the instrument EQ-5D<sup>2</sup>. A health state is then an objectively defined combination of different levels of health attributes, such a combination will be denoted health state profile (HSP). The utility drawn from different health states, on the other hand, has to do with the subjective experience of a health state and is expected to differ between individuals. One could imagine, for example, that a professional athlete would accept some loss of memory in order to keep her mobility while the chess playing scholar would not. This definition's impact on the possibility to measure health is further discussed in chapter 3.

### 1.4.2 Probability

The concept of objective, or true, probability is problematic in a metaphysical sense. It requires a non-deterministic view on causality, a deeper investigation of this subject is however not in the scope of this thesis. An objective probability in this thesis will denote a probability implied by experience, a flipped coin will show heads half of the times because in the past coins have shown heads half of the time. In the medical context this

---

<sup>2</sup> The EuroQoL group is an interdisciplinary group of researchers who have developed an instrument to value different health states. The attributes, or health dimensions, they use are: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. ([www.euroqol.org](http://www.euroqol.org))

may be translated to that a treatment will be successful half of the times because it has been so in the past.

### 1.4.3 Prospect

A prospect is defined as a contract that yields a certain outcome with a certain probability and another outcome with yet another certain probability and so on. The sum of the probabilities always equals one. There is no restriction to how many possible outcomes there can be in a prospect. (Tvesky&Kahneman, 1992)

### 1.4.4 Risk and Uncertainty

That a prospect is risky means that the outcomes in the prospect have a probability that is neither 0 nor 1. Uncertainty means that either the probabilities or the outcomes are uncertain.



## 2. The Theories

With the main concepts defined we are ready to look into the theoretical foundations of the theories that are the subject of this thesis. This chapter contains a definition of notational symbols followed by a brief description of EUT. Then the standard gamble method will be presented followed by a somewhat more thorough presentation of PT. Only the main results of these theories are presented here, for a complete analysis of EUT the reader is recommended to consult an advanced book on microeconomic theory or the original work by v. Neumann, Morgenstern (1944). PT is probably best understood through reading Kahneman&Tversky (1979) and Tvesky&Kahneman (1992).

### 2.1 Notation

The notation will be as follows, capitol letters from the middle of the alphabet, H,I,J..., denote health states and capitol letters from the beginning of the alphabet, A,B,C..., denote prospects. The symbols  $\prec, \succ$  and  $\approx$  denote worse than, better than and equally good as, respectively.  $\succ \approx$  denotes better than or equally good as and  $\prec \approx$  denotes worse than or equally good as. Lower-case letters from the end of the alphabet, x,y,z..., denote outcomes while lower-case letters from the middle of the alphabet, p,q,r..., denote probabilities. U,V denote utility from a prospect while u,v denote utility from outcomes. For illustration: if you are in health state H and choose the prospect A: [x, p; y, q] you will reach health state (H+x) with a probability of p and health state (H+y) with probability q and assuming  $u(H) > u(I)$  iff.  $H > I$ , we can conclude that  $(H+x) \succ (H+y)$  if  $x > y$ . If an outcome is certain then no probability will be stated and if the outcome is zero then no outcome will be stated unless it is crucial for understanding.

## 2.2 Expected Utility Theory

EUT is an axiomatic system that results in a well-known theorem. The widely used theorem has its punch line in the result that the expected utility of a prospect equals the utility of the expected outcomes, weighted by the probability of each outcome.

$$\sum_{i=1}^n p_i u(H + x_i) = U(A) \quad , A : [x_1, p_1; x_2, p_2; \dots; x_n, p_n]$$
 and H is your current asset. Then preferences follow the condition  $A \succ B$  iff.  $U(A) > U(B)$ .

The axiomatic system has been presented in many different ways, e.g. Mas-Collel et.al. (1995) and Hastie (2001). Here I try to give a brief and as far as possible non-technical description of three fundamental axioms.

*Completeness* is the first axiom of EUT. It defines  $\succsim$  as a complete preordering<sup>3</sup>. It entails that  $A \succsim B$  or  $A \precsim B$  or both.

*The axiom of continuity* states that if  $A \succsim B \succsim C$  then there exists a p, between 0 and 1, such that  $B \approx [A, p; C, (1-p)]$ . That this holds for all A,B,C means that no matter how good A is or how bad C is, there will always exist a probability that makes you indifferent between B and a mixture of A and C.

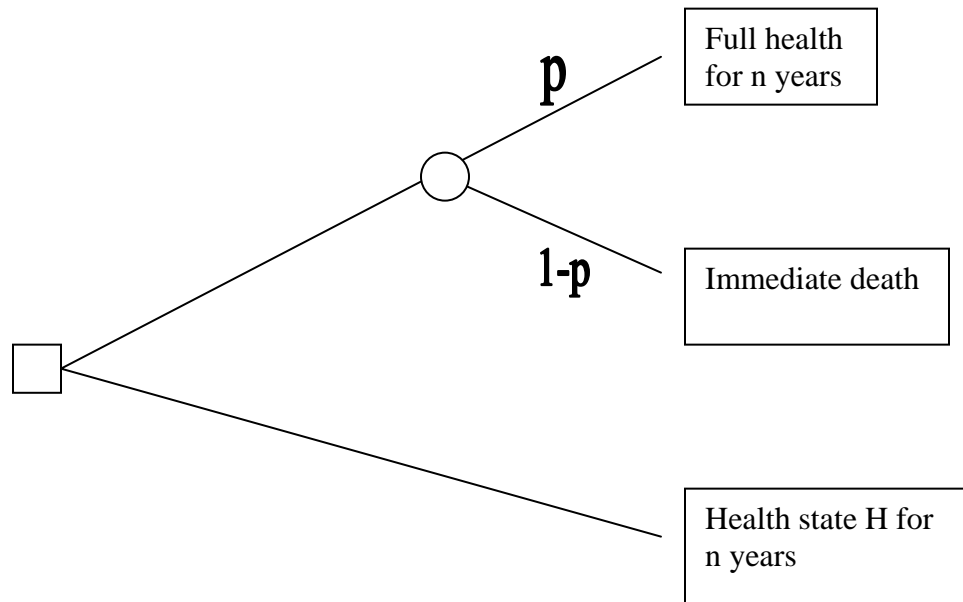
The axiom that seems to be descriptively most problematic and that will be tested later in this thesis is the *axiom of independence*, also called the axiom of substitution, which states that  $[x, p] \succ [y, q]$  iff.  $[[x, p], r; c, (1-r)] \succ [[y, q], r; c, (1-r)]$ . That is to say that if you prefer A to B then the preference will be preserved if you “add the same thing” to both prospects.

---

<sup>3</sup> A complete preordering is a binary relation that is transitive, reflexive, antisymmetric and complete.

## 2.3 Standard Gamble

In the SG respondents are presented with a decision tree (Fig.1) and we want to find out at which probability  $p$  they are indifferent between the two prospects (H) and (Full health,  $p$ ; Immediate death,  $(1-p)$ ). The gamble can be thought of as a risky therapy.



*Figure 1*

If we give full health the quality weight 1 and death the weight 0, then  $p$  can be viewed as the respondent's utility from health state  $H^4$  and used as a quality weight for that health state. (Boardman et al. 2001)

---

<sup>4</sup> According to EUT the respondent's indifference means that  $u(\text{health state } H) = p \cdot u(\text{full health}) + (1-p)u(\text{death}) = p \cdot 1 + (1-p) \cdot 0 = p$ . The existence of such a  $p$  is guaranteed by the axiom of continuity.

## 2.4 The Prospect Theory

The Prospect Theory is a non-conventional<sup>5</sup> theory of decision under risk that includes some psychological factors that are not fully rational in the sense that they are not consistent with the axioms of EUT.

It can be divided into two parts: an *editing phase*, in which a mental representation of the alternatives is constructed and an *evaluation phase*, which consists of assessing the value of the different prospects and choosing the best one.

An important part of the editing phase is choosing a reference point, the outcomes are then viewed as gains or losses compared to this reference point. An intuitively appealing reference point is the status quo; you compare what you get with what you already have, but sometimes it makes sense to assume an aspiration level to be the reference point, you then compare what you get with what you desire.

The evaluation phase can be described by the equation

$$\sum_{i=1}^n \pi(p_i)v(x_i) = V(A) , A : [x_1, p_1; x_2, p_2; \dots; x_n, p_n]$$

, where  $\pi_i$  is called a decision weight and it is a function of  $p_i$ . Important to note here is that  $\pi_i$  does not reflect bad estimations of the probability, or degrees of belief. Instead it reflects attitudes toward probabilities, e.g. that we value certainty in itself. Different functional forms for this weighting process have been proposed (Starmer, 2000) but the important features can perhaps best be shown in a graph (Fig.2). This weighting function was not presented in the original prospect theory but in the “Cumulative Prospect

---

<sup>5</sup> Non-conventional, in the sense that Starmer (2000) uses it, means that the theory can not be reduced to a single utility/value-function.

Theory” published in 1992. In this version the authors allow for different weighting of probabilities for gains and losses and it also covers uncertain situations and allows us to generalize the theory to any number of outcomes within each prospect.

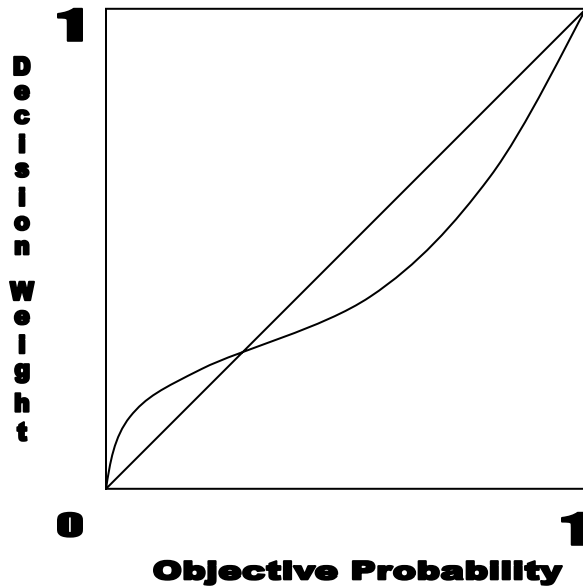


Figure 2

First, we note that  $\pi(0) = 0, \pi(1) = 1$ , decision weights are in the same interval as the objective probabilities. Secondly, the curve is steep close to the endpoints; this illustrates a high sensitivity to the difference between possibility and impossibility and to the difference between certainty and risk.

The other part of the evaluation process,  $v(x_i)$  is to infer utility from the different outcomes. The utility-function proposed by Kahneman&Tversky (1979) is represented with an S-shaped graph (Fig.3).

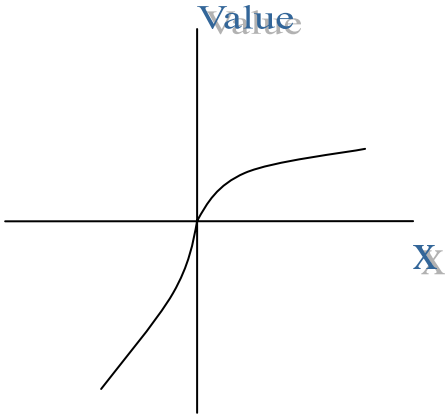


Figure 3

Analytically it is expressed by the value-function

$$V(x) = \begin{cases} x^\alpha & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -\lambda(-x)^\alpha & \text{if } x < 0 \end{cases}$$

, where  $x = 0$  is the reference point (Hastie et al., 2001). Two restrictions define its characteristic features:

$$0 < \alpha < 1$$

, gives concavity when  $x > 0$  and convexity when  $x < 0$ . This implies risk-aversion in gains and risk-seeking behavior in losses. It also entails that the utility-curve is steeper the closer to zero we get<sup>6</sup>, that is to say we are more sensitive to changes close to our reference point.

---

<sup>6</sup> To see this consider the derivatives  $\left[ \frac{d}{dx} x^\alpha = \alpha x^{\alpha-1} = \frac{\alpha}{x^{(1-\alpha)}} \right]$  and

$\left[ \frac{d}{dx} (-\lambda(-x)^\alpha) = \frac{\lambda\alpha}{x^{(1-\alpha)}} \right]$ . They both grow as  $x$  approaches zero.

The second restriction defines *loss-aversion*

$$\lambda > 1$$

, tells us that the utility curve is steeper when  $x < 0$  than when  $x > 0$ . Put differently, a loss hurts more than a gain of the same size gives pleasure.

## 2.5 Cardinality of Health

Here we halt to consider the implications of the fact that there is no cardinal scale for health.

To substitute monetary outcomes for health outcomes might seem like a pretty straightforward approach. However, a problem that has to be solved is that for money a natural cardinal scale exists whereas health is not easily measured on such a scale. There should be no doubt that \$200 is twice as much as \$100 but what health state is twice as much as blindness? Note that there is an important distinction between how much something is and how good something is. That \$200 is twice as much as \$100 does not necessarily entail that we value having \$200 twice as high as that of having \$100, in fact it is often assumed that people have diminishing marginal utility of money. With the discussion of the health concept in mind I propose that there is no universal scale on which health states can be measured. This poses a problem because when we want to observe choices we need to formulate options for the respondents to choose among. My idea to bypass this problem is to get the respondents to first state their valuation of different health states in a setting where risk is not present and then use these valuations in an experiment that resembles the original PT experiments. This way the outcomes are measured on each respondent's subjective scale, in other words the subjective utility from a health state in relation to utility from other health states constitutes the scale on which health is measured.

But if there is no real entity, other than utility, that corresponds to the concept of “one unit of health” then we can not calculate standard economic measures like marginal utility. The utility of being in a health state, if the health scale is constituted by valuations of health states, could of course be nothing else than proportional to the valuation of that health state. In the monetary EUT the utility gained from levels of wealth is calculated, here the utility is given by the respondents definition of health states. This gives us a linear utility of health<sup>7</sup>, and since the only thing in EUT that can explain attitude towards risk is curvature we must conclude that everyone should be risk-neutral with respect to health according to EUT. The PT on the other hand gives us opportunities to explain risk attitudes in two ways; the creation of decision weights and the loss-aversion. The loss-aversion means that if you are currently in health state H, then loosing health to end up in G gives more disutility than going from G to H would give utility. These properties will be further investigated in chapter 5.

## 2.6 The Epistemology of PT and EUT

Before we move on towards developing a method to test if PT outperforms EUT in describing decisions under risk in the domain of health, an effort to anchor these two theories in the basic philosophy of science may be appropriate. To consider the foundations of the theories might be useful in order to understand the problems that arise when we compare them.

EUT is in line with the epistemological tradition of rationalism, that we can reach knowledge by solely thinking. The Prospect Theory is a theory that relies on the epistemological tradition of empirism, that we can gain knowledge by observation. In

---

<sup>7</sup> Remember that in EUT  $\sum_{i=1}^n p_i u(H + x_i) = U(A)$  ,  $A : [x_1, p_1; x_2, p_2; \dots; x_n, p_n]$  , then

$u(H + x_i) = k(H + x_i) \Leftrightarrow k \sum_{i=1}^n p_i (H + x_i) = U(A)$  . This is nothing else than expected final asset, multiplied by an arbitrary constant.



philosophy, supporters of these traditions have argued for centuries regarding questions of how to conduct good science and what a good scientific methodology is.

The EUT is a convincing theory of how we should act if we were completely rational and perfectly informed, even if this is not undisputed (for criticisms see e.g. Allais, 1953 in Hastie et al. 2001, Ellsberg, 1961). It could be argued that the best way to discover the best way of thinking is thinking, observing choices does not tell us whether the process preceding a decision is rational or not. PT on the other hand is a theory that describes actual decisions and the best way to create a theory about what people actually do is, intuitively, to observe what they do. Hence, if we assume that people are not completely rational and informed then there is room for both theories in economics. As Hansen (2006 p.33) puts it when he discusses the relation between neoclassical and behavioral economics: “...*the combination of them [behavioral and neoclassic economics] is better than one of them alone when rendering parts of www [way the world works]*”<sup>8</sup>. The troubles arise when the normative neoclassical EUT is assumed to describe actual behavior.

---

<sup>8</sup> Authors’ explanations within brackets.

# 3. A Method of Investigation

In this chapter the method to test how well PT and EUT describe decisions under risk in the domain of health is developed. First the experiment is briefly described and after that the justifications for it are discussed. In order to add strength to the judgments in this chapter a small pilot study, involving only two respondents, was carried out.

## 3.1 The experiment

The main objective of the experiment is to keep the structure of the prospect theory experiments but exchange the pecuniary outcomes for changes in health. A more detailed description is given in chapter 4.

The experiment is divided into three separate parts:

1. *Learning phase.* The respondents are asked to evaluate five health states, defined by mobility, pains and sight, using a visual analogue scale (VAS)<sup>9</sup> ranging from 0 to 100. The primary objective of this step is to familiarize the respondents with the design of the form for defining health states. It is also ensured that the respondents understand that the scale is meant to be a ratio-scale, that 100 is twice as good as 50 and so on.
2. *Defining health states.* Now the procedure in the learning phase is reversed. The respondents are given the task to define ten health states that fit to given points at a

---

<sup>9</sup> The answer sheet with the VAS is found in the Appendix page 37.

VAS. The main rationale for this step is to make the next part, the gambles, more understandable and concrete.

3. *Gambles*. This is the phase where I try to replicate the prospect theory experiments. The respondents are asked to choose between different prospects containing the health states they have defined. In order to simulate gains it is necessary that the reference point is not full health; the solution to this is to ask the respondents to imagine that they are in another state of health. Another important feature of this experiment is that, in contrast to the original PT experiment, outcomes are presented as final health states and not as changes in health. This is due to the setting where the respondents define their own HSPs.

## 3.2 Learning from Observations

Since the methodology of the behavioral economics is based on empirism it relies on the belief that we can trust our observations. In this section three necessary conditions for eliciting true preferences are identified and their implications on the design of the experiment are discussed.

### 3.2.1 Knowing ones preferences

A fundamental requirement for eliciting true preferences is that respondents know their own preferences; otherwise it would be impossible for them to state them or for the researcher to elicit them from choices made by the respondents

A straightforward solution to this is to assume completeness, as defined in EUT. This is not trivial in all situations; for example it can be hard to imagine how life would be in different health states and hence to know which health state you prefer. However once you have made a choice you have indicated that you know your preferences, in this thesis completeness is simply assumed.

### 3.2.2 Answering Truthfully

Of distinct importance is also that the respondents answer the questions in accordance with their preferences, that they tell the truth. Usually researchers try to encourage the respondents to answer truthfully by compensating them according to their choices; this is called contingent rewards<sup>10</sup>. When investigating health this is not easily done, it is inappropriate to inflict damage upon the respondents as an encouragement to be honest. If they would be monetarily compensated according to their choices there would be a risk that the choices reflect their attitude towards money and risk instead of health and risk. This is a serious dilemma that raises questions about the internal validity of the method presented in this chapter (Corson, 2005). However it is hard to find a solution. The standard procedure is to pay the respondents a fixed sum. Even if they still freely choose how much effort they want to put into answering the questions it could be argued that they make a greater effort if they are paid (Bolton&Ockenfels, 2000).

### 3.2.3 Understanding the questions

To make it possible for the respondents to answer truthfully the questions have to be put in an understandable way. Not understanding what you are asked makes it hard to give a well-grounded reply that mirrors your preferences. However, there is a trade-off between complexity of the questions and precision in the results. To reach a high precision the health states should be defined by as many health attributes as possible and with as many levels of each health attribute as possible. This would give the maximum number of permutations to choose amongst and the best chance that the respondent can define a HSP that fits to a given level of utility. In addition to this, as more health attributes are included we get closer to describing the full concept of health. On the other hand it is important to consider the cognitive effort of the respondents, it is probably harder to comprehend and compare HSPs the more attributes they have. The EuroQoL group uses

---

<sup>10</sup> To illustrate the idea imagine that you want test if someone would accept a 50/50 gamble in which he can either win or loose \$50, then instead of just asking him you offer him to gamble for real. This will ensure that his reply is sincere.

five dimensions with three levels on each giving them  $3^5 = 243$  possible HSPs to fit to a VAS ranging from 0 to 100. However, the task for the respondents in this study is harder than in the EuroQoL group's studies. Not only do they have to rate HSPs on a VAS but they do also have to compare different HSPs to each other. The small pilot study together with introspection supports the assumption that the number of attributes is more crucial to comprehensibility than the number of levels at each health attribute. The HSPs in this method will therefore be defined by three dimensions with seven levels of each which gives  $7^3 = 343$  permutations.

### 3.3 Attributes

The next step in the development of the method is to decide which health attributes to include. One leading principle has been used in doing so, namely that we should, when creating hypothetical health states, only include health attributes that do not have strong correlations with each other. This is desirable not to make the econometric task easier for the researcher but to make the cognitive task easier for the respondent. It might, for example, be hard to grasp what it would be like to have no mobility and at the same time have no anxieties if one believes that decreased mobility entails anxieties. Is it possible to rank or even rate health states one can not imagine? Intuitively it is not. Hence, in order to make use of as many permutations as possible and to avoid a serious threat to the completeness assumption, strongly correlated health attributes should not be included together.

Through common sense reasoning and the mini-pilot three health attributes that are not strongly correlated have been identified. They are mobility, pains and sight<sup>11</sup>.

---

<sup>11</sup> The different levels of each attribute is found in the appendix on page 37.

### 3.4 The Test Procedure

If the linearity of the EUT model is taken literally, then it would be enough to show that respondents are not indifferent between prospects with the same expected payoff to falsify EUT. But that would require an enormous level of trust in that the health state profiles match the value given on the VAS. A defender of EUT could make the very valid objection that the match is not perfect and that that is the reason for the strict preference.

A more convincing test of EUT is to check whether revealed preferences are inconsistent, if a prospect is both strictly better and strictly worse than another at the same time we have a contradiction that is not easily explained.

Testing whether EUT describes the observed choices is the first part of the experiment. If it does not, we want to find out if PT is a better description. The experiment will test five patterns of decisions. Namely:

1. *Common consequence.* The EUT axiom of independence implies that  $A: [x, p; y, q; 0, r] \succ B: [y]$  if and only if  $C: [x, p; 0, r+q] \succ D: [y, 1-q; 0, q]$ <sup>12</sup>. The common consequence effect shows that this implication is not always true. In words, it falsifies the assumption that changing two gambles in the same way regarding *outcomes* will not change the preference order between them.

2. *Common ratio.* The EUT axiom of independence implies that if  $A: [x, p] \succ B: [y, q]$  then  $C: [x, pr] \succ D: [y, qr]$ <sup>13</sup>. The common ratio effect shows that this implication is not always true. In words, it falsifies the assumption that changing

---

<sup>12</sup> We can re write  $B: [y]$  as  $B: [y, (1-q); y, q]$ . If we change  $y$  to 0 in  $q$  percent of the cases, then we get  $C$  and  $D$  as above. The only thing we have changed is the outcome and according to the axiom the preferences should be preserved.

<sup>13</sup> The axiom states that  $[x, p] \succ [y, q]$  iff  $[[x, p]r; c, 1-r] \succ [[y, q]r; c, 1-r]$ . If we set  $c=0$  then  $[x, p] \succ [y, q]$  iff  $[x, pr] \succ [y, qr]$ . The only thing that changes is the probability and therefore the preference order should not change according to the axiom.

two gambles in the same way regarding *probabilities* will not change the preference order between them.

3. *Reflection Effect*. If  $A: [x,p] \succ B: [y,q]$  and  $C: [-y,q] \succ D: [-x,p]$  then there is a reflection effect. The reflection effect suggests that risk attitude is reversed when gains are replaced by losses.

4. *Isolation Effect*. The isolation effect shows that people tend to neglect things that are identical between prospects when comparing them.

5. *Neglecting bonus*. The neglect of bonuses suggests that *changes* in health rather than final health states are the object of evaluation when different prospects are compared.

## 3.5 Validity

In this section the validity of the method as such will be discussed, the survey carried out by the author has some shortcomings that are not related to the method and they will be discussed in the chapter 5. The concept of validity is divided and used as in DePoy, Gitlin (1998).

*Internal validity* is decided by how the question “Do I test what I want to test?” is answered. Firstly, the outcomes are defined in a way that makes us consider the possible problem that respondents look at the numbers rather than the health states<sup>14</sup>. When asking them, however, they said they compared the health states and not the numbers. I also heard respondents reasoning in line with “I gain some mobility but the probability is lower” which suggests that this is not a problem. However there exists technical solutions to this, e.g. in a computerized experiment this could be taken care of. Secondly, it is not

---

<sup>14</sup> See answer sheet on page 37

obvious that the respondents are used to working with the concept of probability; an illustrative aid<sup>15</sup> may help understanding probabilities.

*External Validity* deals with to what extent the results of a study are possible to generalize to the study population. The capacity to generalize in this context is the degree to which the answers in the study coincide with the choices that would have been made in a real life situation. Firstly, reactivity is problematic. An interviewer effect is difficult to avoid in an experimental situation where an interview leader is present, this puts us in a dilemma. We could either create a questionnaire to fill out without guidance and risk that the questions are incorrectly interpreted, a loss of internal validity, or apply a method like the present and risk influencing the respondents. I suggest the latter alternative with, if possible, a well trained interviewer. Secondly, the realism of the experimental situation raises questions about the capacity to generalize over the results. As mentioned before contingent rewards are not possible in this setting, therefore this is a seemingly unsolvable problem. Furthermore, it could be difficult to imagine health states and, perhaps, much to ask from someone to simultaneously imagine that they are in another health state and choose between hypothetical treatments.

*Construct Validity* refers to how well constructs and concepts are operationalized. With the method presented in this thesis the concept of health is in effect reduced to only three health attributes, they do not fully capture health defined as a composite good. But these attributes could be varied across studies, that they are only three is primarily a way to safeguard the internal validity of the method by minimizing the cognitive complexity of the task to compare different health states. The concepts risk and probability on the other hand seems to be not at all problematic to operationalize.

*Reliability* addresses the stability of the design. If repeating the interviews under the same conditions would yield the same results, then we the design would be reliable. This is of course possible to test. In this thesis a limited test of the reliability of VAS is performed,

---

<sup>15</sup> See Appendix, page 36, for an example.



it showed that some degree of arbitrariness is present when defining health states to fit a certain value on the VAS. The result is further discussed in the next chapter.

# 4. Experiment

In this chapter the results from the experiment performed by the author are presented. Moreover, the testing procedure is illustrated and shown in its application.

## 4.1 Sample

The study was performed on seven male respondents, who are a relatively homogenous group. They were recruited from the author's personal social network. The interviews lasted between half an hour and an hour. Some descriptive statistics are shown in table 1 below. The typical respondent is a healthy young man who is in the end phase of his studies.

*Table 1 Descriptive statistics for the sample*

	Average	St. Deviation	Range
Age	25.1	0.4	25-26
Education <sup>16</sup>	193.6	53.1	90-260
Health status	95.1	4.9	90-100

---

<sup>16</sup> Education is measured in Swedish university credits. 40 credits correspond to one year's workload.

## 4.2 The Health State Profiles

The different levels within each health dimension are assumed to be ranked on an ordinal scale. The worst level is put first<sup>17</sup>. Given this assumption, no inconsistencies were found in the way the respondents rated the health state profiles. Nobody rated a dominated health state higher than the dominating health state. It is also worth to note that there was a vast variance in the way the respondents defined the health states, e.g. a health state that was valued as 70 by one respondent was valued as 50 by another. This supports the methodological principle that the respondents should define their own health states instead of using an average from previous surveys.

Ten HSPs were defined corresponding to the values 10, 20, 30, 49, 50, 51, 70, 80, 85 and 90 on the VAS. It turned out that most of the respondents were not able to discriminate between 49, 50 and 51. This might be due to a lack of appropriate combinations of the attribute levels. An improvement of the method in this respect can be achieved either by adding more levels or by making the difference between the health states larger. The reason for the minimal difference is that those health states are used to simulate a lottery and an insurance, and it is interesting to observe choices when it is “cheap” to participate.

## 4.3 Validity of the Experiment

In addition to the discussion on validity in the method-chapter this survey carries some problems of its own. Firstly, I know the respondents personally which may strengthen the interviewer effect. Secondly, the sample is too homogenous to claim much external validity. The homogeneity together with the small number of participants makes any statistical analysis of the background variables useless. If the experiment is performed as a full-size survey with a random sample it is possible to test if background variables have

---

<sup>17</sup> See the form in the appendix on page 37.

any correlations with the answers and hence judge the external validity of this experiment.

As mentioned earlier, no inconsistencies were found in the HSPs, this indicates that they have some degree of reliability. However, a small reliability test concerning the HSPs was performed. One week after the first interview the respondents were once again asked to define the health state corresponding to 50 on the VAS. The result of this test was that three respondents defined exactly the same HSP as they did the first time, two respondents defined a health state that was better in one dimension but worse in another and two respondents defined HSPs that were worse in all dimensions, i.e. a HSP that was strictly dominated by the one they defined a week earlier. There are at least two plausible interpretations of this result. One could argue that the shift in valuations does not matter because the experiment was carried out directly after the first valuation and that those valuations must have been valid then, and that the results a week later reflects a real change in valuation. On the other hand one could also argue that the shift in valuation shows that the HSPs are not a stable reflection of the respondent's real attitudes. In the latter case the HSP corresponding to 50 on the VAS should be interpreted as a health state that gives approximately 50 units of utility, give or take a few. In the conclusions the latter interpretation will be made.

## 4.4 The Gambles

With exception of the bonus neglect test, the respondents were asked to assume that they were in the health state they had just defined as 50. In the analysis it is assumed that this health state is adopted as reference point. The outcomes are changes relative to that health state. For example, the prospect [30] is presented as reaching the health state 80 for sure. Practically, they are shown the form on which they have defined the health state that corresponds to 80 on the VAS together with a circle diagram representing the probability; in this case the circle would be filled.

## 4.5 Results from the Experiment

In the following presentation of the results the absolute number of respondents that made a specific choice is within parenthesis. A prospect presented as [35, 0.33; 30, 0.66] shall be interpreted as 33% chance to gain 35 units of health, 66% chance to gain 30 units of health and 1% risk to not improve at all. The result of each problem is presented together with a brief analysis.

### *Common consequence effect*

Problem 1	A: [35, 0.33; 30, 0.66]	(3)	B: [30]	(4)
Problem 2	A: [35, 0.33]	(6)	B: [30, 0.34]	(1)

This is a clear violation of the EUT axiom of independence; if B is preferred to A in 1 then B should also be preferred to A in 2, the only difference between 1 and 2 is that a 66% probability of gaining 30 has been replaced by a 66% probability of gaining nothing. Three respondents made the modal choice, the modal choice means choosing A in one problem and B in the other, and thereby violated the axiom.

### *Common ratio effect*

Problem 3	A: [40, 0.8]	(2)	B: [30]	(5)
Problem 4	A: [40, 0.2]	(5)	B: [30, 0.25]	(2)

Again the EUT axiom of independence is violated, the only difference between 3 and 4 is that the probabilities are four times larger in 3. According to EUT this should not change the preference order. Five respondents made the modal choice.

Problem 5    A: [40, 0.45]                    (2)    B: [20, 0.9]    (5)  
 Problem 6    A: [40, 0.01]                    (7)    B: [20, 0.02]    (0)

This is another example of the common ratio effect. Again five participants made the modal choice and thereby violated the axiom of independence.

*Reflection Effect*

These problems correspond to problem 3-6 but with losses instead of gains.

Problem 7    A: [-40, 0.01]                    (1)    B: [-20, 0.02]    (6)  
 Problem 8    A: [-40, 0.45]                    (3)    B: [-20, 0.9]    (4)

Only two respondents showed the modal preference.

Problem 9    A: [-40, 0.2]                    (1)    B: [-30, 0.25]    (6)  
 Problem 10    A: [-40, 0.8]                    (5)    B: [-30]            (2)

Five modal choices were counted.

*Table 2 Illustration of the reflection effect*

Problem 3: [40, 0.8] < [30] (2)    (5)	Problem 10: [-40, 0.8] > [-30] (5)    (2)
Problem 4: [40, 0.2] > [30, 0.25] (5)    (2)	Problem 9: [-40, 0.2] < [-30, 0.25] (1)    (6)
Problem 5: [40, 0.45] < [20, 0.9] (2)    (5)	Problem 8: [-40, 0.45] < [-20, 0.9] (3)    (4)
Problem 6: [40, 0.01] > [20, 0.02] (7)    (0)	Problem 7: [-40, 0.01] < [-20, 0.02] (1)    (6)

In the table you can see that, except for the third row, preferences are reversed when gains are replaced by equally-sized losses. This is the *reflection effect*. As seen in the first



*-You have just been treated so that your current health state is 90, now you need an additional treatment.*

Problem 13    A: [-20, 0.5]                      (3)    B: [-10]                      (4)

Problem 12 and 13 resemble each other. Choosing A gives is a 50/50 gamble between ending up in health state 70 and 90 while B puts you in health state 80 for sure. The only difference is that in 12 you gain health and in 13 you loose health. Two respondents made the modal choice indicating that the change rather than the final health state is evaluated. However, the other five opted for either A in both or B in both problems. A possible explanation for this is that the outcomes in this experiment are presented as final health states and not as gains or losses, this requires a rapid shift of reference point for the respondent to perceive the difference between the two problems. In the original PT experiment the respondents had “their own” reference point and the outcomes were presented as gains or losses, assumingly that made the task easier for them. This might be the reason that this result differs from the PT. Another plausible explanation is that the final health states are evaluated instead of gains and losses.

*Lottery/Insurance*

Problem 14    A: [50, 0.02]                      (7)    B: [1]                      (0)  
Problem 15    A: [-50, 0.02]                      (0)    B: [-1]                      (7)

In four cases the health states 49, 50, 51 were defined as the same health state. Three respondents were able to discriminate between the health states but still made the modal choice. This result suggests an overweighting of small probabilities.

## 4.6 Conclusions from the Experiment



The only statistical analysis that will be undertaken here is to note that it is improbable that the replies are random. If the choices were made randomly we would expect an even distribution between the alternatives. But here  $64/91=0.70$  (excluding the lottery/insurance result) of the choices were made in accordance with PT. The probability of this or a more extreme distribution to occur, given random choices, is very slight. A test of the hypothesis that choices are random returns  $Z=3.82$  implying  $p < 0.01$ , hence we can confidently refute the hypothesis.

Another conclusion from this experiment is that EUT was heavily violated. In all cases EUT was violated by at least two respondents and in most cases the majority violated EUT. Hence we have clear support for the statement that EUT is not a good descriptive theory for decisions under risk regarding health<sup>18</sup>.

Now we turn to the PT. Which of the PT features are found? First of all, there are indications of a fundamental difference between this method and the original PT. In this setting it was not clear whether the respondents evaluated the changes in health or final health states. Further experiments are needed to come to a clear conclusion regarding this.

Most pronounced of the PT characteristics is the probability weighting. We observed a certainty effect that is illustrated in figure 2 by the steepness close to the northeast corner,  $\pi(p) < p$ , put different probabilities close to certainty are underweighted. We also observed a low sensitivity to differences between small probabilities. This suggests a concave curvature close to southwest corner. Furthermore, the lottery/insurance problems suggest overweighting of small probabilities in accordance with figure 2.

No unambiguous risk-seeking behaviour in losses was found, the risk-seeking attitude suggested by problem 10 was reversed in problem 8 (see table 2). Risk-aversion in gains seems more robust, as revealed by problem 3 and problem 5. Regarding loss-aversion no

---

<sup>18</sup> Note that the common consequence and common ratio effect do not rely on that the HSPs reflect any particular utility level.

reliable conclusion can be drawn. To come to a conclusion regarding loss-aversion we should use problems of the form A:  $[x, 0.5; -x, 0.5]$  B:[0], this was realized by the author after the experiment was carried out and is therefore still left to investigate.

In the experiment we could observe choices that did not support PT, e.g. not everyone chose the certain outcome in problem 3. However, that does not necessarily give us reasons to abandon PT. There are two reasons for that, first of all differences between individual preferences are expected, the theory has to describe some kind of average behaviour. Secondly, even if choices are not predicted by PT they can still be incorporated in PT in some cases. Consider for example problem 1, A:  $[35, 0.33; 30, 0.66]$ ; B: [30]. Choosing A implies that  $\pi(0.33)v(35) + \pi(0.66)v(30) > v(30)$ . This can be true even if we assume a tendency to underweight the probabilities in A. The reason for choosing A could be that the effect of the decision weights does not outweigh the difference in value.

A final conclusion from the experiment is that if it is carried out in a larger scale more conclusions can probably be drawn,

## 5. Conclusion

The general conclusion of this thesis is that it is in principle possible to design experiments that test decisions under risk in the domain of health. The major difference compared to experiments with monetary outcomes is that health has no measurable objective dimension. This fact requires that every respondent constructs his own subjective scale for health states. Furthermore, investigations in the domain of health entail some validity problems that are not easily solved. Perhaps, the most severe of them is that the respondents in such an investigation may have troubles comprehending the outcomes, health states, used in the experiment. In the developed method efforts have been made to minimize these problems. The small scale experiment that was carried out suggests that those efforts were satisfactory.

The experiment also suggested that EUT, in particular the axiom of independence, is systematically violated also in the health context. Even if the experiment has problems with its validity, e.g. a small sample, it seems rather clear that EUT is not a suitable theory for describing decisions under risk in the domain of health.

The question whether PT is a better description is not easily answered based only on the small experiment in this thesis. Nevertheless, some support for the basic principles of the creation of decision weights was found.

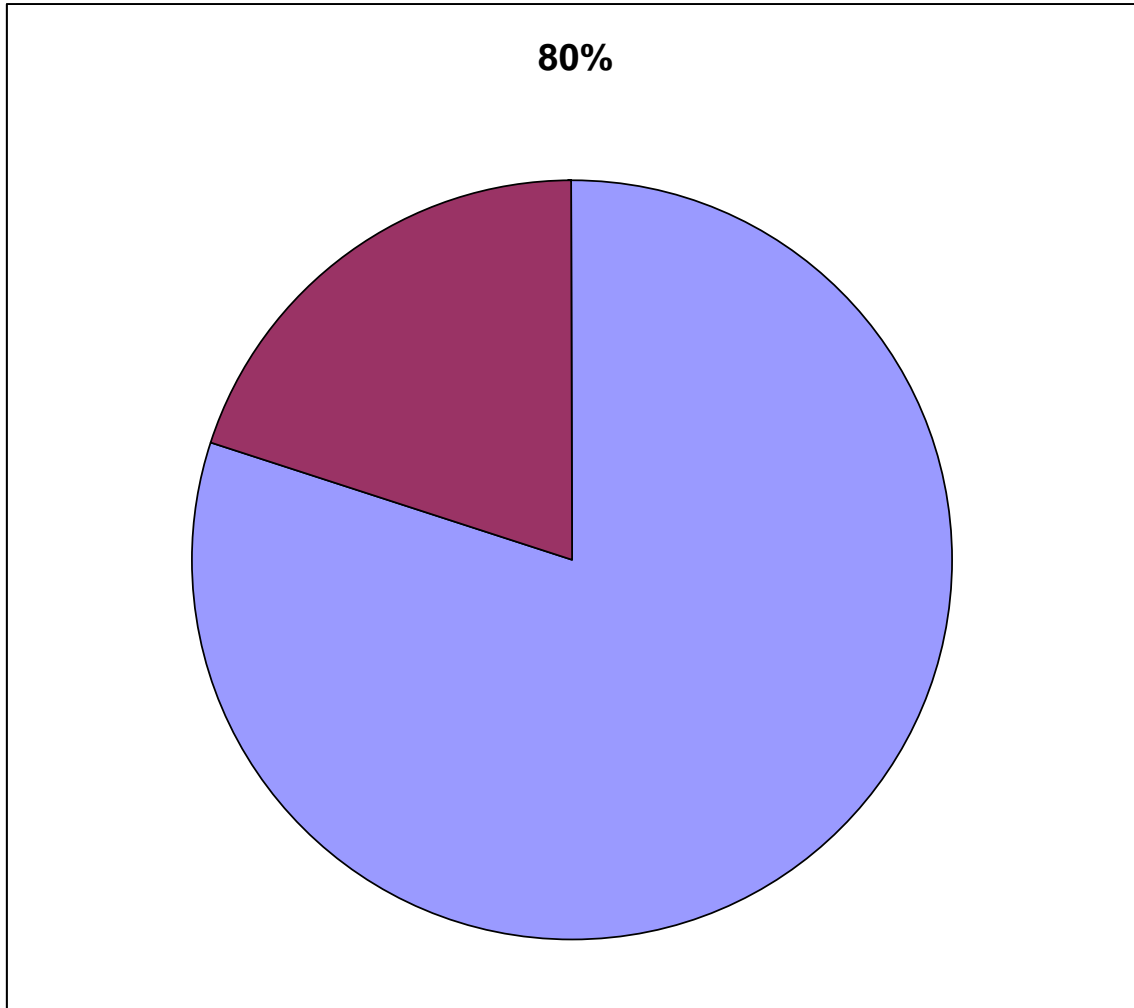
## 5.1 Discussion

In the introduction I claimed that, to show that PT is a better descriptive theory than EUT, it is not enough to show that using PT improves the internal consistency of a survey and that PT has to be tested itself. Therefore, a critical question that needs to be discussed is: In what way will the method developed in this thesis give more decisive arguments than e.g. Bleichrodt et.al. (1999) and Oliver (2003) have already provided? Firstly, this method gives the opportunity to formulate the problems in almost any way one wants; one is not restricted to the frames of the standard gamble. This makes it possible to test smaller fractions of the theories separately. For example, this method makes it possible to show the existence of the common consequence effect and to analyze the reasons for the creation of decision weights; this is not easily done in a standard gamble setting. Secondly, the previous analyses of decisions under risk in the domain of health have primarily been concentrated to the duration of life and not to the quality of life. Bleichrodt et.al. (1999), however, involve quality of life in a section of their study and they find that applying a probability weighting function improves internal consistency. Hence, their conclusion is supported by the results from the experiment in this thesis.

Finally, the method can be used in other settings. First, it would be interesting to perform experiments both with more dramatic and with less dramatic changes in health. One could either change the outcomes in the problems or change the levels of each health attribute to accomplish this. My guess is that decisions are made in different ways depending on how much is at stake. Second, decisions regarding other attributes than mobility, pains and sight could be investigated with this method. Such investigations would, if they give the same results, increase the construct validity of the conclusions; we would come closer to describing decisions regarding the whole concept of health.

## 6. Appendix

*Illustration of probability, aid used in the experiment.*



Skalan till höger finns för att underlätta för dig att värdera olika hälsotillstånd. Värdet 100 på skalan motsvarar bästa tänkbara hälsotillstånd och 0 på skalan motsvarar sämsta tänkbara hälsotillstånd. Hälsotillstånd definieras genom olika nivåer av smärta, mobilitet och syn.

Värdet 50 ska ses som ett hälsotillstånd hälften så bra som 100.

Värdet 25 ska ses som 25 % så bra som 100 o.s.v.

Markera vilken kombination av egenskaper som svarar mot det på skalan markerade värdet.

### *Smärta*

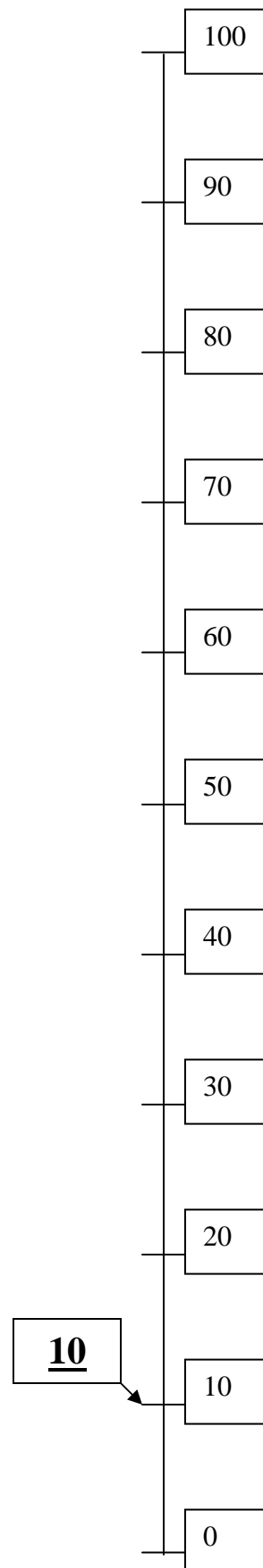
- Ständig svår värk
- Ständig lindrig värk
- Ständig ömhet
- Tillfällig svår värk
- Tillfällig lindrig värk
- Tillfällig ömhet
- Fri från smärta

### *Mobilitet*

- Förlamad.
- Rullstolsbunden.
- Behöver kryckor för att gå.
- Kan inte springa, haltar.
- Kan inte springa, går normalt.
- Svårigheter att springa.
- Full mobilitet.

### *Syn*

- Blindhet
- Kraftigt nedsatt syn, ser endast konturer och ljus.
- Blind på ett öga, inget djupseende.
- Vindögdhet
- Färgblindhet
- Nedsatt syn, åtgärdas med glasögon eller linser.
- Perfekt syn.



# 7. References

1. Bleichrodt H, Pinto J.L, Wakker P. Making Descriptive Use of Prospect Theory to Improve the Prescriptive Use of Expected Utility, *Management Science* 47: 1498-1514 (2001)
2. Bleichrodt H, v. Rijn J, Johannesson M. Probability Weighting and Utility Curvature in QALY – based Decision Making, *Journal of Mathematical Psychology* 43: 230-260 (1999)
3. Boardman et al. *Cost-Benefit Analysis Concepts and Practice*, Prentice Hall (2001)
4. Bolton G E, Ockenfels A. ERC: A Theory of Equity, Reciprocity and Competition, *The American Economic Review* 1: 166-193 (2000)
5. Camerer C F. *Prospect Theory in the Wild: Evidence From the Wild*. Social Science Working Paper 1037 California Institute of Technology (1998)
6. Croson R. The Method of Experimental Economis. *International Negotiation* 10: 131-148 (2005)
7. DePoy E, Gitlin L N. *Introduction to Research*, Mobsy inc. (1998)
8. Doctor J N et al. A NEW and More Robust Test of QALYs, *Journal of Health Economics* 23: 353-367 (2004)
9. Ellsberg D. Risk, Ambiguity and the Savage Axioms, *Quarterly Journal of Economics* 75: 643-669 (1961)
10. Folland S, Goodman A C, Stano M. *The Economics of Health and Health Care*, Prentice Hall (2004)
11. Hansen F. *Explorations in Behavioral Economis: Realism ontology and Experiments*. Lund Economic Studies 133 (2006)
12. Hastie R and Dawes DM. *Rational Choice in an Uncertain World*. SAGE (2001)
13. Tversky A and Kahneman D. Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty* 5: 297-323 (1992)
14. Kahneman D and Tversky A. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47: 263-292 (1979)
15. Machina M J. Choice Under Uncertainty: Problems Solved and Unsolved, *Economic Perspectives* 1 121-154 (1984)
16. Mas-Colell A et.al. *Microeconomic Theory*, Oxford University Press, (1995)
17. Oliver A. The Internal Consistency of the Standard Gamble: Tests After Adjusting for Prospect Theory, *Journal of Health Economics* 22: 659-674 (2003)
18. Starmer C. Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk. *Journal of Economic Literature* 38: 332-382 (2000)
19. Von Neumann J and Morgenstern O. *The Theory of Games and Economic Behavior*. Princeton, N.J. Princeton University Press. (1944)
20. [www.euroqol.org](http://www.euroqol.org) 2007-01-18