Msc in Health Economics, Policy and Law

Institute of Health Policy and Management

Erasmus University Rotterdam

Master Thesis

QALY MODEL: Test of Utility Independence when health varies over time

Name: Ilias Pyrnokokis

Student number: 361799

Professor: Han Bleichrodt

Table of contents

ABSTRACT

Objectives: In recent years, economic evaluations of health care interventions are widely undertaken in order to inform policy makers in the health care sector about the cost-effectiveness of the old and new technologies. Regarding the benefits of these technologies, the QALY model is the most used measure for utility. It is an index that combines quality and quantity of health after the elicitation of people's preferences over the interventions, which are going to be compared. For QALY to be a valid instrument when elicit people preferences, some fundamental principles should hold. One of them is Utility Independence. There is empirical evidence, which supports QALY in regards to utility independence for chronic health states. However, there is limited evidence when health varies over time.

Method: This study consists of two parts. In the first part, three tests of utility independence are conducted by using the standard gamble method to elicit people's utilities over various health profiles. The health states that were used to form the profiles were taken from EQ-5D-5L. After the elicitation of utilities we did two statistical tests, one paired t-test and one Wilcoxon test in respect to each utility independence test. The second part of our study consists of a questionnaire in order to investigate sequencing effects based on people's general preferences over mild and severe health states.

Results: In our first part of the study where utility independence was tested, statistical analysis showed that utility independence holds, since there were no statistically significant differences between the utilities for each test. Regarding the qualitative research through the questionnaire we created, we explored some sequences effects, such as adaptation to mild disability. We believe that this happened due to limitation in the procedure of elicitation.

Discussion: Utility Independence is supported through our study; however, we should consider some limitations that don't allow us to conclude over the validity of the instrument. A relative small sample (n=30), the method we elicited utilities (SG), the difficulty of our sample to involve in such a complex experiment and the puzzling results of our qualitative part where some sequences effects were found, lead us to the conclusion that more research should be done for the QALY in order to be certain of its validity, especially in the field of sequencing effects on people's preferences.

INTRODUCTION

Challenges of our times such as scarce resources and increase of population by health improvements have led health care sector, among others sectors, to the problem of being unable to get sufficient funding to a level where demand will be met (Brazier, 2007). Modern methods and technology development may be beneficial for the people, but they demand resources that are either not available or cost too much, especially today where we are experiencing recession all over the world, and countries together with private entities are struggling to contain their costs and keep providing adequate services. As a result of this, resources should be managed and distributed with caution and taking into account the final goal of maximizing well-being of the society. Questions such as of who will pay, who will be benefit from health care use, what and where we should focus our research on, which services will be provided as a priority, are of high importance and need valid answers. That is why economic evaluation of health care interventions is highly and widely used. It is a very important component in the formation of the policies and the actions that will be taken by decision makers in health care sector. More specifically, the goal of economic assessments is to be a tool that will support and develop rationality by comparing modern and old interventions through analysis of their costs and effects. This is done in order to give this added value regarding the knowledge over the real value of these interventions, which will be applied to the population (Drummond et al., 2005). This means that maximization of the total health benefits can be achieved if after the efficacy and the effectiveness part in the assessment of an intervention, a relative cost-effectiveness is applied when competed (Bobinac et al., 2011).

Economic evaluation consists of three different types: Cost-effectiveness analysis, cost-utility analysis and cost-benefit analysis (Brazier, 2007). What they have in common is the feature of cost in their analysis, but what is different and distinguishes each type is the way they measure and evaluate benefits of health care. Moreover, economic evaluation can be categorized through the different perspectives that are used when conducting it, from a health care perspective to a societal one (Bobinac, 2011). In reality, official institutions that are giving the guidelines for economic evaluation in health care suggest that a societal perspective should be taken in order to include a wider range of costs and benefits, regardless who is using them or gaining utility through consumption of health care (Gold et al., 1996)

As we mention above, the conducting of economic evaluation is differentiated in terms of how the benefits are measured with each method, since the costs are always valued in monetary units (Drummond et al, 2005). Regarding benefits, in cost-benefit analysis the consequences are valued as the cost do, in monetary units, while in cost-effectiveness analysis are valued in natural units, such as life-years gained, disability-days saved, points of blood pressure reduction etc. Last but not least, in cost utility analysis, it is used a broader measure of the effects after the implementation of a health care program and that is utility. Utility is meant to explain people's preferences over health outcomes, such as a health state or even a health profile that is formed by constituent health states over time.

There are quite few different models to measure utility, such as Disability Adjusted Life Years (DALY), Healthy Years Equivalents (HYE) and Quality Adjusted Life Years (QALY). In this paper, it is in our concern the last model that is the most popular when conducting economic evaluation for health care interventions.

QALY

Since QALY is the most widely used measure for the utility, the focus in research is to investigate if this model calculates utilities with validity, more specifically, if the utility obtained fully represents people's real preferences. Although the sensitivity of the issue examined and since there is a big discussion of what "real" preferences actually are, we will refer to this debate later in the paper.

QALYs are calculated by trying to quantify the quality of life or to rephrase it, trying to incorporate both quality and quantity of life in a single measure index (Weinstein et al., 2009). This effort to adjust years of life to the quality of life which is spent needs caution, especially when it is a vital part of assessing health care projects and determine the funding of health care sector and its implication on people's health by health care received. The QALY approach considers a health profile, which is formed by constituent health states and proposes that the utilities obtained by each health state can be added and retrieves the utility of this health profile with respect to time, that is the application of discounting when they are used in cost-utility analysis. Suppose that we have four health states, q1, q2, q3 and q4 respectively, which form a health profile Q. The use of QALY model assumes that we can obtain with validity the utility of profile Q by adding the utilities of its constituent health states q1, q2, q3 and q4 with respect to people's preferences for time (Spencer and

Robinson, 2007). It is obvious that this makes it an additive model and its calculation can be represented as:

$$U(Q) = w1 \times U(q1) + w2 \times U(q2) + w3 \times U(q3) + w4 \times U(q4) ,$$

where w is the discount factor for time and U expresses a utility function. The QALY model has advantages (Bleichrodt and Filko, 2008) when used in cost-utility analysis i.e., it is computed easily and it is a measure that can be understood easily by policy-makers and doctors, which is something really important in a multi-disciplinary sector like health care where specialists of different skills interact, and communication in this cooperative process is of high importance. Its disadvantage lies on the fact that we cannot be sure if the model represents people's preferences since its validity depends on some restrictive assumptions, which don't hold always when people express their preferences over a, desire or not, health state.

Literature review on QALY measurement distinguishes two situations. One is when health states are chronic which implies that quality of life is constant over time. In this situation the assumptions made are that these chronic health states are preferred to death and expected utility holds (Pliskin et al., 1980). The latter is assumed in both situations Expected utility is a theory of decision-making under risk and it is logical to make such an assumption because we evaluate the elicitation of people preferences under risk and risk is incorporated when we take decision for health care interventions where the results of their implications are uncertain. Moreover, in order for the QALY model to be valid when health is constant, three additional assumptions have to be made (Pliskin et al., 1980). The first is mutual utility independence, which implies that utility of quality of life is independent of life duration and life duration is utility independent of quality of life. The second assumption imposed is constant proportional tradeoffs, which holds if the proportion of the life years that someone is going to give up for an improvement in his quality of life doesn't depend on the absolute number of these life years. The last assumption is risk neutrality with respect to life duration and that is for a specific level of quality, each prospect is indifferent to its expected value. Furthermore, Bleichrodt et al. (1997) showed that by assuming zero condition- which was first introduced by Miyamoto and Eraker (1988) and refers to the fact that when life duration is zero, health qualities are valued equally- the only assumption that is needed to be examined in order the QALY model to hold is risk neutrality with respect to life duration. In several studies it is showed that QALY model is valid for chronic health states although there are other studies which express criticism (Bleichrodt and Pinot-Prades , 2006).

QALY WHEN HEALTH VARIES OVER TIME

However, the most realistic case is the one where health varies over time (Bleichrodt & Filko, 2008) and this is the case that concerns this paper. In order QALY model to be valid when health varies over time and different utilities are attached to its constituent health states of a health profile, the strongest assumption that we should make is additive independence, which implies that utility of a certain health state in the health profile doesn't depend on any other health state included in the health profile, otherwise, it holds if the preferences between risky treatments depend only upon the marginal rather that the joint probability distributions of the different health qualities (Bleichrodt & Quiggin, 1997), Keeney and Raiffa, 1976). Furthermore, beside additive independence, it is important to imply symmetry with respect to each time period when we value health quality. Imposing symmetry means that the constituent time periods that the health qualities are valued must be similar. Bleichrodt and Quiggin (1997) confirmed that by showing that QALY model is valid under expected utility when additive independence and symmetry hold. Empirical evidence exists on testing additive independence (Spencer 2003) and showed that this central assumption doesn't hold, but due to ambiguous results through the limitations in data collection process, the QALY model cannot conclusively be rejected. Other studies (Bleichrodt & johannesson 2001, Chapman 1996, Cairns & van der Pol 1997) showed that symmetry also doesn't hold because it is inconsistent with discounting.

Although the assumption of additive independence is a strong one (Spencer & Robinson 2007), the QALY model can still be valid if the assumption of utility independence holds. Utility independence implies that within a certain health profile, the value we attach in each health state is independent of the health state that is before or after it. This weaker assumption, if it holds, is capable of supporting a valid QALY model even if we need additional adjustments, such as to estimate the weights of different health states in a health profile. However, as a result of this, we will end up with a model which is not a conventional one. Three studies of Spencer and Robinson (2007), Bleichrodt and Filko (2008) and Spencer and Robinson (2004) support QALY model after testing utility independence.

Together with the assumption of additive independence and utility independence, Keeney and Raifa (1976) include also a case of riskless choice-in contrary with the first two ones which are dealing with risky choices- and that is preferential independence. This assumption holds if preferences between profiles, which include a common health state at a specific period of time, do not depend on the health state in that period. Treadwell (1998) showed that preferential independence holds in 36 out of the 42 tests that were conducted.

Moreover, Guerrero and Herrero (2005) relaxed utility independence assumption and thus, they provided a semi-separable approach of QALY model. What they actually do is to distinguish between initial utility independence and final utility independence. By that separation, it is implied that a person's preferences over prospects, for example between two risky treatments, which incorporate future health quality and have the same past, do not depend on what is this past. This model requires utility independence only for the initial periods of an individual's life and thus, this situation is mentioned as initial utility independence. Unfortunately, when health varies over time, there is no direct evidence on initial utility independence.

Another situation, which is worth mentioning when health varies over time, is to characterize QALY when expected utility is violated. In that case, additive independence cannot hold. However, it has been suggested that additive independence can be replaced by generalized marginality (Bleichrodt & Quiggin, 1997) through a general utility model and additive representation of a health profile still exists as in the case of additive independence under expected utility. The empirical evidence (Bleichrodt & Filko, 2008) regarding the test of generalized marginality showed that it could not be rejected in aggregate level, thus the QALY model is valid when it is applied in economic evaluation in health care, but that was not the case at the individual level where it is needed a more general QALY model in order to be consistent with people's preferences.

SEQUENCES OF HEALTH STATES

In the first chapter, we explained the concept of the conventional QALY model; particularly the focus was based on the implementation of the model when health varies over time. In this case, health profiles are formed by constituent health states in order to be assessed based on people's intertemporal preferences over this health profile. If these preferences satisfy "additive independence" (Guerrero & Herrero, 2005) under expected utility, then utility over any given health profile is additive, which satisfies the principles needed for the QALY model to hold. Also, the model assumes utility independence, which means that the conditional preferences for each part of the health profile's attributes should be independent of its complement (Guerrero & Herrero, 2005).

Furthermore, since we discuss about preferences over different health profiles, it is needed to explain at this point the concept of utility, which is the level of satisfaction people enjoy through the choices they make, hence utility represents their preferences. The concept of

utility can be interpreted in two ways, either as experienced utility or decision utility (Dolan & Kahneman, 2008). The former was firstly introduced by philosopher Bentham (1789/1948), where utility is defined in hedonic terms and measures the pleasure people feel or the deviation of pain in that pleasure. The hedonic view of utility (Dolan & Kahneman, 2008) is related to the concept of happiness. Although this concept has its origins back to the 19th century, experienced utility was abounded due to the critic that it cannot represent people choices and it is not measurable (Loewenstein & Ubel, 2007).

Instead of the "great happiness" principle, modern decision theory is choice-based, which means that utility is inferred from revealed preferences, thus decision utility (Varey & Kahneman, 1992). This concept is closer to the rationalistic point of view in behavioral economics, where people choose based on what they want, rather than what they enjoy as the experienced utility implies. As a result of this concept of "wantability" (Fisher, 1918) and the dominance of decision utility after the 20th century in the application of economic analysis regarding people preferences, neoclassical welfare economics assume that people are rational, fully-informed and try to maximize utility when choices are made (Dolan & Kahneman, 2008).

Although decision utility has dominated for many years, the rationality of human beings is questioned in many aspects of real-life choices they make. Do people choose the best for themselves when they make decisions? Do they really know what maximizes their utility? Literature review provides strong evidence against this rational character of humans, especially when they make choices under risk. Research on the elicitation of utilities when choices are made shows that people do not behave in an effort to maximize their utility, hence their actions don't satisfy the two fundamental requirements that rationality implies, consistency and coherence (Tversky & Kahneman, 1981). The violations of these two characteristics suggest that psychological principles govern people's perspective when dealing with decision problems. Particularly, the way that the problem is framed will determine the choice that will be made, challenging the maximization of the utility derived from that choice. Without a proper sequential context (Loewenstein & Prelec, 1993), choices over different sequences of outcomes may not reflect rationality. For example, when a decision maker has to choose between two health profiles, he may choose the worst one, which actually gives him a lower quality of life, if the frame of the question is not appropriate to make him realize what is best for him.

Choices under risk are based on the theory of expected utility, where the utility of a risky prospect is equal to the expected utility of its outcome (Tversky & Kahneman, 1981). Also the QALY model measures utilities through elicitation of people's preferences based on expected utility. Given these facts as well as the additive nature of the model, the violations of rationality and decision utility must be addressed in order to investigate possible drawbacks of the model, which is applied to the assessment of health interventions and has been a determinant in the allocation of scarce resources.

Firstly, research on framing effects over sequences of outcomes illustrates that the frame that a person will follow depends more on the way that the given problem needing to be solved is formulated than the actual beliefs and norms the person has derived from his cultural background (Tversky & Kahneman, 1981). As a result of this, a problem can be seen in multiple ways, as well as its solution, making the outcome and its utility different, depending on the various perspectives of its formulation. Under risk, responses to the problem can be different even for the same person if, for example, the same outcome will be presented to him as a gain instead of a loss. The deviation from his initial perspective could lead him to this inconsistency mentioned before. Within the concept of framing effect, we can distinguish two more phenomena of choices, the certainty effect and the pseudo certainty effect (Tversky & Kahneman, 1981). The former is violating the rationality on choices due to the diversity of choices that have to be made under risk and as a result of these choices people tend to prefer the safest outcome although it may not be the most rewarding. The latter violates a stronger assumption that implies that preferences are independent of the formulation of the given problem.

More evidence of inconsistency regarding the stable character of choices over sequences of outcomes that neoclassical theory suggests, is the research of Loewenstein and Prelec (1993), where it is shown that people have the tendency to prefer improving sequences of outcomes, meaning that utility should be improved over time and this is a violation of additive separability. For example, people prefer improving wages profiles over time rather than decreasing regardless if the total outcome is the same in both situations (Loewestein & Prelec, 1993). This observation has multiple explanations. Firstly, it could be due to the "happy ending" phenomenon, where people tend to prefer ending up with a gain as a perception of satisfaction. Moreover, this is in line with Kahneman's theory (Kahneman et al, 1993) which suggests that final stages are over weighted in a sequence of outcomes, making them a major determinant in the overall assessment of this set of outcomes.
Documentation for such preferences exists also in the research of Ross and Simonson (1991), where well ending has been illustrated in a sequence of outcomes.

Secondly, preferences for improvement over time can be explained by savoring and dread (Loewenstein & Prelec, 1993), where people wait for the best outcome until the end and remove dread by getting rid of bad outcomes in the early stages.

Thirdly and most importantly, the contrast effect is another major determinant of preferences improving over time. This refers to the behavior of people comparing current situations and the feelings derived from them with relevant situations faced in the past or that will be faced in the future. This reference point, intuitively revealed to people's actions when they make choices, is based on the two concepts of adaptation and loss aversion (Kahneman et al, 1993). By adaptation, we refer to the psychological mechanisms that people use to adapt in stimuli over time and each time assess a new one based on this adaptation (Kahneman et al, 1993). For example, a patient with a chronic cancer condition,

assuming 5 years of already being ill, will adapt to this situation and evaluate his utility higher than he would do if he wouldn't have cancer, due to adaptation to his condition. Similar empirical evidence as the above was observed in the study of Tversky and Kahneman (1981). Since this paper is dealing with the concept of QALY model and taking into consideration its impact in policy field as it is used for the evaluation of health and health care interventions, it is important to state at this point that this phenomenon of adaptation may result to an unequal allocation of resources in health care, since adaptation may not reveal the real needs for services. However, no adaptation doesn't guarantee optimal allocation of resources by its own. Using the same example with the patient with cancer, if he does adapt to his condition and expresses less pain than cancer as a disease is actually causing, he may get less priority in treatment funding than other groups of patients with a different disease, although both groups may have the same level of severity (Dolan & Kahneman, 2008). However, someone can claim that if people with cancer in general adapt easily, then this less priority is justified, but such problems must be dealt with cautious in such a sensitive field where human lives are involved.

The second attribute of contrast effect is loss aversion. Within this concept, people are giving more weight to losses than to gains judged from a neutral reference point (Tversky & Kahneman 1979). Hence, once again people may prefer a different sequence of outcomes if they think of a gain than of a loss and vice versa, although the total utility is the same in both situations. That means a hypothetical value function (appendix) in regards to diminishing sensitivity will be S-shaped, meaning that it will be convex below the reference point while concave above it (Tversky & Kahneman, 1981). This greater sensitivity in losses than in gains can be illustrated by an example of the empirical evidence where people's dissatisfaction with the loss of an amount of money is greater than the satisfaction of gaining the same amount (Tversky & Kahneman, 1981). The concept of loss aversion is explained by prospect theory. Prospect theory is applied to correct these inconsistencies in people's preferences and since traditional QALY model is based on expected utility, there are concerns of its validity when loss aversion is observed. Moreover, if we assume that the most recent outcome is the determinant of the reference point in regards to loss aversion, this strengthens the effect of improving sequences (Ross & Simonson, 1991) on people's preferences.

Another mechanism that may violate additivity and independence between the various levels in a sequence of outcomes, promoting the improvement of the latter over time, is the recency effect (Miller & Campbell, 1959) in regards to retrospective judgments of such a set of outcomes. As a result of this retrospective perspective in the assessment of a sequence of outcomes (Varey & Kahneman, 1992), people give more weights to the final levels of the stream.

Continuing with the investigation of retrospective evaluations, it is important to introduce at this point the monotonicity criterion, where it is stated that adding pain in a set of outcomes,

should increase the disutility in the total outcome (Kahneman et al, 1993). This rule can be accompanied with the notion of non-discrimination, meaning that there is no rationale to evaluate an experience by putting more weight than another. However, real evidence shows violation of these concepts. In one investigation (Redelmeier & Kahneman, 1996), it is illustrated that people don't care about the duration of the event but they are focusing on the peak and end of this event. Taking account other similar studies (Varey & Kahneman, 1992), the common finding is that people put more weight on final and worst moments, and they are unconcerned with the duration for an experience they have (Kahneman et al, 1993). Considering the retrospective evaluations of an experience and the pattern of peak-and-end that has been observed, we have to mention that duration may be not the biggest determinant, but that does not mean that it plays no role. Evidence (Hsee & Abelson, 1991) shows that velocity may be a factor in the assessment process of an experience. Kahneman et al (1993) explains that this may be caused by the power of the initial moments of an event over the memories of this event. Last but not least, the integration rule which states that the weights of all the moments in a sequence of outcomes should be equal, has been found violated in a research of Varey and Kahneman (1992), suggesting that people try to simplify the information given for the solution of the problem, which means that these simplified heuristics that follow are in contrast with the concept of decision utility.

Evidence was presented in this chapter which supports an alternative welfare criterion based on experienced utility. In contrast with decision utility, where the assumption rests on the fact that people are rational regarding their preferences and the goal of their choices under risk is to maximize utility, it was illustrated that people do not act like that in many situations and the frame of the problem is a big determinant of the final outcome, meaning that different perspectives of an experience can give different utility levels of it. Adaptation in specific situations that affect judgment of the event, the chronological order in a sequence of outcomes, positive time discounting in a series of an event, loss aversion, peak-and-end pattern in retrospective evaluations, and either recency or primacy effects (the levels in the beginning of a sequence matter in the latter situation), all suggest that there are psychological mechanisms that don't allow people to think in a normative way as expected utility suggests, hence validity of models in decision-making which are based on this rational notion of behavior, such as the traditional QALY, are challenged when applied.

However, critique is also applied for the concept of experienced utility. As we illustrated above, public policies can be misled by the effect of hedonic adaptation (Loewenstein & Ubel, 2007) by not taking into account important values that are not related with happiness or happiness isn't reflected properly. The lack of incorporating non-hedonic elements of an experienced event will give biased results in regards to people subjective happiness, or to put it differently, happiness doesn't reflect everything that makes a life valuable. Moreover, the critique yields also to the fact that there is no gold standard for happiness and as a result of this, there are no valid measures for evaluating something that vague. But even if

we define subjective happiness and we can measure it with consistency, there will always be a debate whether public policy should focus only on maximizing happiness, as it is expressed through experienced utility, or focus on other criteria for maximizing society's well-being. The famous philosophical question of John Stuart Mill whether it is better to be a dissatisfied human being or a happy pig, is still lacking an answer. Do we choose paternalistic policies that are most likely driven by the concept of experienced utility, or do we choose libertarian approaches giving the freedom of choice, assuming people are rational human beings as decision utility does? Loewenstein and Ubel (2007) suggest that a policy's various alternatives shouldn't be characterized as either black or white, and propose informed decision utility as a solution for public policy, where the notion of decision utility can be applied if the people make choices after getting well informed over this multi-dynamic task in the field of decision-making. We think that this proposal closely matches to the concept of Libertarian Paternalism, expressed by Sunstein and Thaler (2003), where freedom of choice is of great importance, but guidance and provision of information to the public is necessary for the optimal results as we try to maximize society's well-being.

METHODOLOGY

As already mentioned, in this paper we tested whether in the QALY model utility independence holds or not, when health varies over time. Keeney and Raiffa have suggested (Keeney & Raiffa, 1976) that, for this kind of test, a paired gamble should be applied to ensure validity. However we used the simple standard gamble technic for reasons of simplicity and after deciding to follow former relevant studies' methodology. Utility independence suggests that when people are taking into account risky choices in their life, their preferences over these choices should be independent of the element's severity that is common throughout these choices when compared (Spencer & Robinson, 2004). Assuming that we have four different health states (elements) A, B, C and D which can be combined in order to form various health profiles, such as ABC, BCC, CCB etc, depending on the combination we choose. What utility independence means is that if a respondent is indifferent between a certain profile BCC and a p% chance of profile BAA and 1-p% chance of profile BDD, then they should be also indifferent between a certain profile CCC and a p% chance of profile CAA and 1-p% of profile CDD.

We created different set of profiles, which all of them, having in common the duration of 45 years, divided in three health states of 15 years duration. The health states were derived by EQ-5D instrument (The EuroQol Group, 1990), which describes health states of five

dimensions: mobility, self-care, usual activities (working, studying etc.), pain/discomfort and anxiety/depression. Each dimension has 5 level of severity: no problems, slight problems, moderate problems, severe problems and inability/extreme problems (**APPENDIX A**). Although the instrument can derive many possible health states, in our test we considered only four health states for reason of simplicity and understanding to our respondents. Firstly, normal health state, which explains no problems in all dimensions, was color-coded with pink and denoted as A. Respectively, mild health state with moderate problems, green and as B, severe health state with severe problems, red and as C, and finally Death as D. For example, a 45 years-old health profile ACC denotes 15 years in normal health state, followed by 30 years in severe health state.

Initially, we explained these four heath states to our sample and operated a practice standard gamble question in order to get them used to the experiment before they were presented with the real ones. Such a technic was also followed in the research of Spencer (2003). Hence, the practice question was to evaluate a gamble consists of the best outcome AAA (45 years in normal health) and the worst outcome DDD (Death 45 years earlier), to the certain outcome of AAD (30 years of normal health followed by death) and by presenting to them possible probabilities of success, which means leading them to the best outcome, they were asked in each case, whether they prefer the gamble, the certain outcome or they are indifferent between these two. Most important was to understand the concept of indifference, and that was not only because it was fundamental for the real tests, but also because it was noticed that people had a hard time realizing what indifference means in such situation. We will revisit that in the discussion part.

After the description of the health states and the practice gamble question, we introduced them to the scenarios (parts of the tests) (**APPENDIX B**). We made it clear that all the scenarios have only two elements in common. Firstly, they had to consider in every scenario that they are the victim in a car accident at the age of 20, and that the health profiles that followed after the accident have duration of 45 years, divided to three different health states of 15 years duration.

Then, they were directly presented with 3 tests of utility independence; each consisting3 of two SG questions, A and B, creating 6 scenarios in total. The scenarios were mixed, meaning that, for example, the 1A question of test 1, was the first question they had, but 1B of the same test were presented to them as the fifth question in the row. Because the purpose of utility independence is to give the same probability to both questions of each

test, we address the questions separately in order to avoid patterns of familiarity with the situations.

First test explores the impact of changing the health state in the first period (first health state) of life from mild to severe. Second test explores the change in the last period (third health state) from mild to severe and the third one explores the change in the duration of two health states (first and second) in the row, checking the difference from mild health state to the normal one for more than one period. The tests can be seen in table 1.

After obtaining the utilities, we conducted a second part of the research by presenting to our respondents eight questions in order to investigate their general preferences over sequences of health states. Hence, they were asked whether they prefer to delay outcomes of ill health or they prefer to experience them in the beginning of their life and get rid off them quickly. Moreover, they were asked if they think that mild and severe health state is more tolerable or less tolerable through time. This was done in order to see if people think they can adapt to ill health situations (mild and severe) as the time passes or not.

We used a sample of 30 people with average age of 24, 5 years. There were 19 males and 11 females. We excluded health economists or people who are familiar with our concept. Also we excluded patients or people who have faced a serious car accident. We chose people who are educated. In regards to the latter, our rationale was based on the fact that education promotes a rational way of thinking, but we are aware of the fact that well-educated people can have imprecise preferences as well. Finally, our hypothesis is that, regarding the QALY model when health varies over time, Utility Independence holds, hence we expect that people will give the same indifference probabilities respectively to the two questions of each test.

RESULTS

After obtaining the probabilities of indifference through the six standard gamble questions in order to analyze three tests of utility independence when health varies over time, we

used statistical analysis with paired t-tests and Wilcoxon tests through SAS software. In both situations we used a significance level of α=0.05.

The null hypothesis in our analysis is that there should be no statistically significant difference between the indifference probabilities in respect to each test. The paired t-tests results support the null hypothesis in all tests. As can be observed in the **APPENDIX C, Table 1** where the results of paired t-tests are illustrated, p value in all situations (0.10 – 0.76 – 0.58) is higher than a=0.05 after checking for equality of variances, and we can conclude that utility independence holds for all the three tests: change in initial health, change in final and the change in duration. The Wilcoxon test, **APPENDIX C, Table1** suggests the same for a=0.05, where p values are 0.07 for test 1, 0.87 for test 2, and 0.085 for test 3 respectively. Indeed, the data shows that prior to the analysis people did not deviate a lot from response to response respectively to each test. Obviously, there were some extreme differences, but this could be due to the procedure of the elicitation of the indifference probabilities. For example, in the first test, one person gave a probability of indifference really low compared to the others in regards to certainty of severe health state, which was perhaps because he felt that a severe health state is worse than it is in reality. This is shown by the fact that the subject insisted many times that severe health state is close to death, although the difference was explained to him through the description of the health state. It was the same person who responded the same way when the severe health state was just in the end of his life (third period of 15 years duration, from 50-65 years old) after the guarantee of normal health for most of his life. Overall, the responses were consistent with the concept of utility independence for each pair, supported also through the statistical analysis.

At this point, we present the results of the second study, where we elicit people's general preferences over the sequences of health states. The results are presented in **APPENDIX C, Table 3**. This study was almost the same as the study conducted by Spencer and Robinson (2007) and we can see that the results followed a similar pattern to those in their paper.

Since we tested for sequencing effects, we would expect people of our sample to disagree with questions 1 and 3 and agree with questions 2 and 4 (**Table 3**) if the hypothesis was that no sequence effects exist through people preferences. Although the majority of our sample responded by supporting this hypothesis, in the case of the question where "prefer mild disability in the beginning of the sequence" was imposed, more people (15 vs. 11)

agreed with the statement, but this can be explained due to the fact that they want to get rid of this health state soon.

Another purpose of this study is to investigate behaviors of adaptation either to mild health state or to the severe one. If people don't adapt in these health states, we should expect them to response negatively in questions 5,6,7,8. The results for this part of the qualitative research suggest that sequencing matters, due to the fact that their responses are based on the severity of health state. We can observe that they agree that mild disability becomes more tolerable with time (27 vs. 3), hence that they adapt to this health state through time. Equally, they agree that severe disability is less tolerable through time (18 vs. 8). Hence, severity of their health state is important for them in decision-making.

DISCUSSION

We conducted three tests of utility independence accompanied with qualitative research through a questionnaire that checked for respondents' preferences over sequences of health states. The QALY model suggests that preferences satisfy the condition of utility independence; hence we can derive total utilities from constituent parts. Our study suggests the same since our analysis showed that utility independence holds, which is one of the fundamental principles for the model of QALY to be valid. However, caveats of the study and concerns arisen through the results and the procedure of the experiment have to be mentioned.

First of all, we did observe sequencing effects in the second part of our study, where people took into account severity of health by stating that they do adapt to mild disability and expressing the preference that severe disability is indeed less tolerable through time. Those results support the theory of adaptation (Kahneman et al, 1993) in the case of the mild disability, and revealed sequencing in regards to severe disability. Also, we found support for the theory of Loewenstein and Prelec (1993), since we saw a pattern of behavior that follows the notion of dread. People showed a strong preference for mild disability in the beginning of the sequence, meaning that they want to remove dread by getting rid of it. However, this wasn't the case with severe disability.

Regarding the method used for the tests, we have to take into account that the probabilities of indifference elicited may be higher from what they should be in the reality, and that is because we used the standard gamble method, which empirical evidence (Bleichrodt & Johannesson, 1997) has shown that, in general, utilities derived from standard gamble are higher comparing with other methods, such as TTO. This was not tested in our study. However, potential upward biases wouldn't affect the utility independence tests, but we believe it needs to be mentioned.

Moreover, we have to point out that the sample was relatively small (**APPENDIC C, Table 2**). Similar studies which test the QALY model used bigger samples, even twice ours. However, we do believe that the selection of the sample was done with caution in order to avoid biases. As it was mentioned before, we exclude people who had a car accident or are patients to avoid subjectivity in responses. We try to address the questions to people who can represent the average taxpayer who is responsible for funding, even indirectly, the health care system.

Furthermore, it could be difficult for the respondents to take into account all the components of such an experiment. People had to think about different health states, duration of life, an unpleasant car accident scenario, and probabilities. As a result of this complex situation, preferences may be imprecise (Butler & Loomes, 2007). We didn't and couldn't test for the level of this imprecision but we should take into account the fact that many respondents expressed difficulty in giving a number. There was a subject that was excluded due to the fact that he felt miserable when asked to respond to a car accident. Although the components of the experiment were explained before beginning with questions, many people expressed their difficulty to understand the meaning of "indifference" when they were asked for the indifference probability between the certain outcome and the gamble. We did our best to explain it, but we have to consider biases due to this fact and suggest caution in future research. Also, many people were anxious when severe disability was expressed, either to the utility independence tests or the questionnaire. Although the health states were described, they kept thinking of severe disability as almost dead when they decided on probabilities, which was not the case obviously. Maybe the color of red we used for denoting this health state was too aggressive for them or we didn't explain it adequately. Regarding the representation of the experiment to the sample, we used pen and paper with health states being color-coded. Although colors help in the distinguish of the different health states, it would be better to conduct the experiment on a computer screen, since visualization of the questions through graphs (Chapman, 2000) supports more concentration and stimuli for the subjects. In general, dynamic methods are more effective than static ones (Spencer & Robinson, 2007).

Moreover, since we randomized the standard gamble questions and each pair was not questioned simultaneously, we believe that our research is robust in terms of avoiding people responding strategically out of the simplicity of how they formulate their answers, and thus supports utility independence.

We also think that the qualitative part of our research gives an insight to sequences effects, such as adaptation, and as suggested by other researchers (Spencer, 2003), more weight should be given to this direction when testing preferences of people in order to capture more elements of violation in expected utility, such as loss aversion. The empirical evidence in regards to the sequences of outcomes shows that people deviate systematically of what is considered as rational behavior; hence we suggest more weight to qualitative research as more effective in capturing and explaining any deviations that may exist.

To conclude, we find support for utility independence, but take into consideration all the above elements explained as concerns through the research we conducted. Preferences of people should be elicited with caution, especially when they will be responsible for the funding of such an important aspect of our lives as health and health care. More research should be done in regards to the QALY model and its fundamental principles in order to conclude whether it is a valid measurement and a trustworthy component in health care decision-making. Hence we recommend its use in health policy, but only as a supplementary tool for policy makers and not as the only truth in decision making process.

Appendix A. Description of health states used in the experiment.

NORMAL HEALTH STATE A	
Mobility	**No** problems in walking about
Self-care	**No** problems with self-care
Usual Activities	**No** problems doing my usual activities (work, study, housework etc)
Pain/Discomfort	**No** pain or discomfort
Anxiety/Depression	**Not** anxious or depressed

MILD HEALTH STATE B	
Mobility	**Moderate** problems in walking about
Self-care	**Moderate** problems with self-care
Usual Activities	**Moderate** problems doing my usual activities (work, study, housework etc)
Pain/Discomfort	**Moderate** pain or discomfort
Anxiety/Depression	**Moderately** anxious or depressed

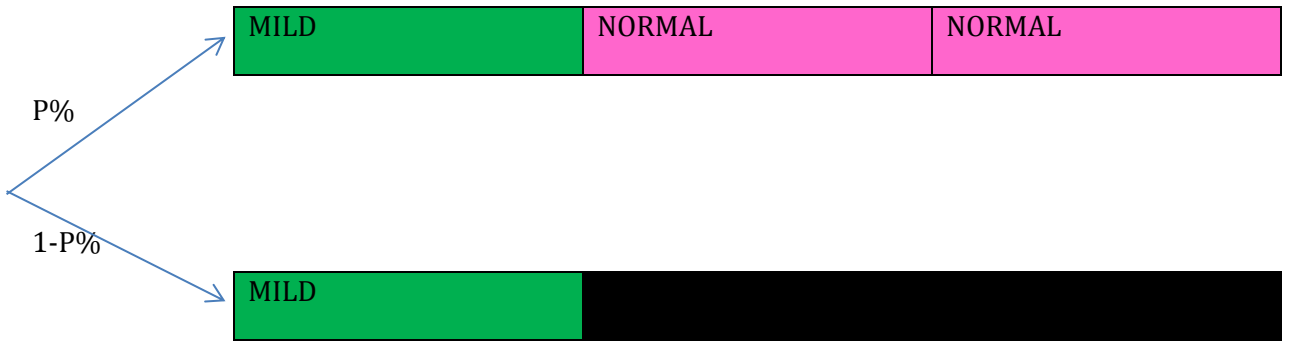
SEVERE HEALTH STATE C	
Mobility	**Severe** problems in walking about
Self-care	**Severe** problems with self-care
Usual Activities	**Severe** problems doing my usual activities (work, study, housework etc)
Pain/Discomfort	**Severe** pain or discomfort
Anxiety/Depression	**Severely** anxious or depressed

And DEATH, color coded black.

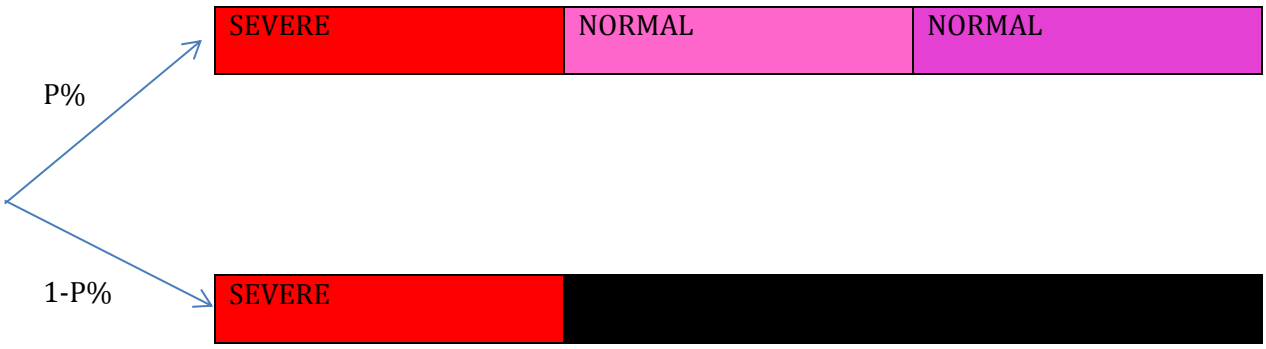
APPENDIX B

TEST 1 – UTILITY INDEPENDENCE

EACH STATE HAS DURATION OF 15 YEARS.

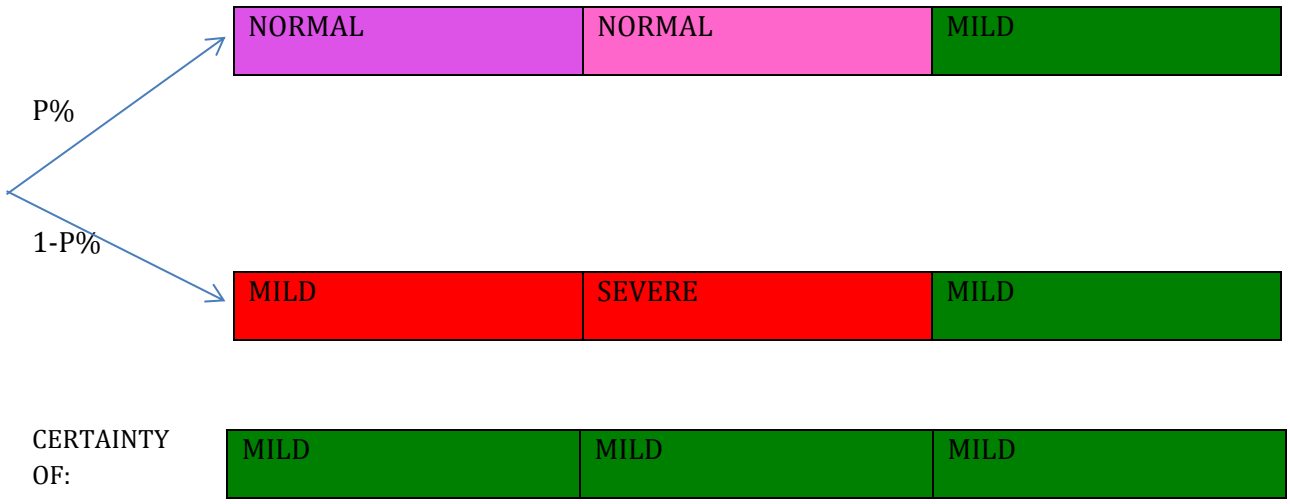


EACH STATE HAS DURATION OF 15 YEARS.

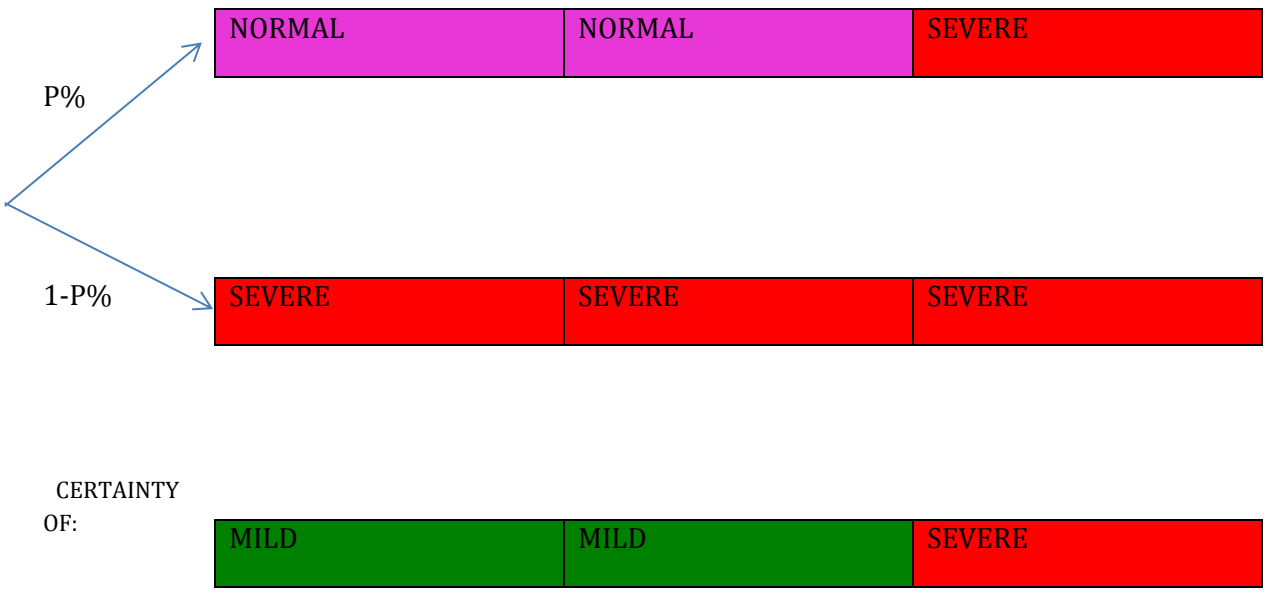


TEST 2 – UTILITY INDEPENDENCE

EACH STATE HAS DURATION OF 15 YEARS.

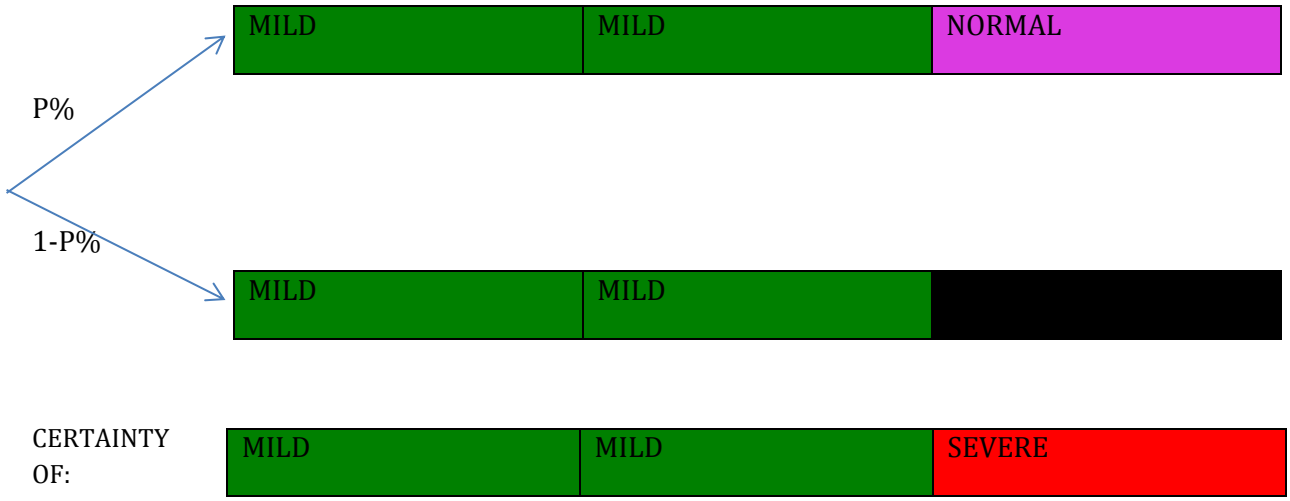


EACH STATE HAS DURATION OF 15 YEARS.

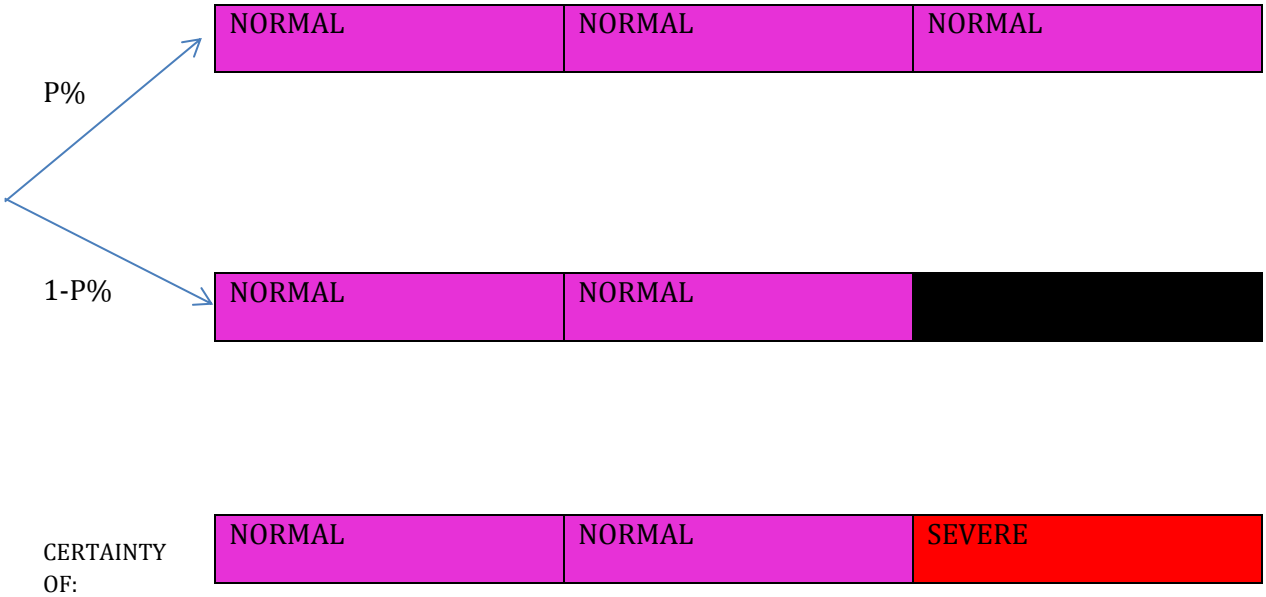


TEST 3 – UTILITY INDEPENDENCE

EACH STATE HAS DURATION OF 15 YEARS.



EACH STATE HAS DURATION OF 15 YEARS.



APPENDIX C

Table 1. P values after statistical analysis through SAS.

TEST UI	P value – Paired t-test	P value – Wilcoxon test
TEST1	0.10	0.07
TEST2	0.76	0.87
TEST3	0.58	0.085

Table 2. Indifference probabilities obtained from the sample.

SAMPLE	1A	1B	2A	2B	3A	3B
1.	73	25	92	91	50	15
2.	94	94	80	70	99	99
3.	66	21	95	75	21	7
4.	30	30	80	90	90	90
5.	75	70	90	91	90	90
6.	89	95	90	85	90	75
7.	80	90	95	95	80	84
8.	64	54	94	98	89	98

9.	70	74	59	85	70	78
10.	55	59	84	70	80	79
11.	69	69	85	87	88	88
12.	55	58	98	85	78	77
13.	60	48	58	63	63	54
14.	78	63	73	83	78	78
15.	65	44	70	68	78	87
16.	57	48	87	90	80	83
17.	50	45	78	83	89	69
18.	45	50	55	78	60	55
19.	60	30	88	91	80	65
20.	69	58	74	82	82	82
21.	60	50	95	80	90	93
22.	63	51	89	80	93	83
23.	85	78	64	93	71	82
24.	72	73	81	84	72	76
25.	69	70	88	86	80	84
26.	52	58	81	73	80	80
27.	55	55	80	84	87	82
28.	98	93	93	85	81	88
29.	37	45	90	95	60	60
30.	70	40	90	80	75	65

Table 3. Results of study 2

	QUESTION	AGREE	DISAGREE	UNSURE
1.	Prefer severe disability in the beginning of the sequence	8	20	2
2.	Prefer severe disability in the end of the sequence	21	7	2
3.	Prefer mild disability in the beginning of the sequence	15	11	4
4.	Prefer mild disability in the end of the sequence	18	9	3
5.	Mild disability becomes less tolerable through time	3	26	1
6.	Mild disability becomes more tolerable through time	27	3	0
7.	severe disability becomes less tolerable through time	18	8	4
8.	severe disability becomes more tolerable through time	6	23	1

References

Bentham, J. (1879) An Introduction to the Principles of Morals and Legislation. Clarendon Press.

Bleichrodt, H. and J. Quiggin (1997) 'Characterizing QALYs Under a General Rank Dependent Utility Model', Journal of Risk and Uncertainty 15(2): 151-165.

Bleichrodt, H., P. Wakker and M. Johannesson (1997) 'Characterizing QALYs by Risk Neutrality', Journal of Risk and Uncertainty 15(2): 107-114.

Bleichrodt, H. and M. Filko (2008) 'New Tests of QALYs when Health Varies Over Time', Journal of health economics 27(5): 1237-1249.

Bleichrodt, H. and J.L. Pinto (2005) 'The Validity of Qalys Under Non-Expected Utility*', The Economic Journal 115(503): 533-550.

Bobinac, A., van Exel, N Job A, F.F. Rutten and W.B. Brouwer (2011) 'Health Effects in Significant Others Separating Family and Care-Giving Effects', Medical Decision Making 31(2): 292-298.

Brazier, J. (2007) Measuring and Valuing Health Benefits for Economic Evaluation. Oxford University Press.

Butler, D.J. and G.C. Loomes (2007) 'Imprecision as an Account of the Preference Reversal Phenomenon', The American Economic Review : 277-297.

Cairns, J.A. and van Der Pol, Marjon M (1997) 'Saving Future Lives. A Comparison of Three Discounting Models', Health Economics 6(4): 341-350.

Chapman, G.B. (2000) 'Preferences for Improving and Declining Sequences of Health Outcomes', Journal of Behavioral Decision Making 13(2): 203-218.

Chapman, G.B. (1996) 'Temporal Discounting and Utility for Health and Money.', Journal of Experimental Psychology: Learning, Memory, and Cognition 22(3): 771.

Dolan, P. and D. Kahneman (2008) 'Interpretations of Utility and their Implications for the Valuation of Health*', The Economic Journal 118(525): 215-234.

Drummond, M., Mark J. Sculpher, G.W. Torrance, B. O'Brien and Greg L. Stoddart (2005) Methods for the Economic Evaluation of Health Care Programmes. Oxford: Oxford University Press.

Fisher, I. (1918) 'Is "Utility" the most Suitable Term for the Concept it is used to Denote?', The American Economic Review 8(2): 335-337.

Fisher, I. (1918) 'Is "Utility" the most Suitable Term for the Concept it is used to Denote?', The American Economic Review 8(2): 335-337.

Gold, M.R. (1996) Cost-Effectiveness in Health and Medicine. Oxford University Press.

Guerrero, A.M. and C. Herrero (2005) 'A Semi-Separable Utility Function for Health Profiles', Journal of health economics 24(1): 33-54.

Hsee, C.K., R.P. Abelson and P. Salovey (1991) 'The Relative Weighting of Position and Velocity in Satisfaction', Psychological Science 2(4): 263-266.

Johannesson, M. and P. Johansson (1997) 'Is the Valuation of a QALY Gained Independent of Age? some Empirical Evidence', Journal of health economics 16(5): 589-599.

Kahneman, D. and A. Tversky (1979) 'Prospect Theory: An Analysis of Decision Under Risk', Econometrica: Journal of the Econometric Society : 263-291.

Kahneman, D., B.L. Fredrickson, C.A. Schreiber and D.A. Redelmeier (1993) 'When More Pain is Preferred to Less: Adding a Better End', Psychological Science 4(6): 401-405.

Keeney, R.L., H. Raiffa and D.W. Rajala (1979) 'Decisions with Multiple Objectives: Preferences and Value Trade-Offs', Systems, Man and Cybernetics, IEEE Transactions on 9(7): 403-403.

Loewenstein, G.F. and D. Prelec (1993) 'Preferences for Sequences of Outcomes.', Psychological review 100(1): 91.

Miller, N. and D.T. Campbell (1959) 'Recency and Primacy in Persuasion as a Function of the Timing of Speeches and Measurements.', The Journal of Abnormal and Social Psychology 59(1): 1.

Miyamoto, J.M. and S.A. Eraker (1988) 'A Multiplicative Model of the Utility of Survival Duration and Health Quality.', Journal of Experimental Psychology: General 117(1): 3.

Pliskin, J.S., D.S. Shepard and M.C. Weinstein (1980) 'Utility Functions for Life Years and Health Status', Operations research 28(1): 206-224.

Redelmeier, D.A. and D. Kahneman (1996) 'Patients' Memories of Painful Medical Treatments: Real-Time and Retrospective Evaluations of Two Minimally Invasive Procedures', Pain 66(1): 3 <last_page> 8.

Ross, W.T. and I. Simonson (1991) 'Evaluations of Pairs of Experiences: A Preference for Happy Endings', Journal of Behavioral Decision Making 4(4): 273-282.

Smith, D.M., G. Loewenstein, P. Rozin, R.L. Sherriff and P.A. Ubel (2007) 'Sensitivity to Disgust, Stigma, and Adjustment to Life with a Colostomy', Journal of Research in Personality 41(4): 787-803.

Spencer, A. (2003) 'A Test of the QALY Model when Health Varies Over Time', Social science & medicine 57(9): 1697-1706.

Spencer, A. and A. Robinson (2007) 'Tests of Utility Independence when Health Varies Over Time', Journal of health economics 26(5): 1003-1013.

Sunstein, C.R. and R.H. Thaler (2003) 'Libertarian Paternalism is Not an Oxymoron', The University of Chicago Law Review : 1159-1202.

The, E.G. (1990) 'EuroQol-a New Facility for the Measurement of Health-Related Quality of Life', Health Policy 16(3): 199-208.

Treadwell, J.R. (1998) 'Tests of Preferential Independence in the QALY Model', Medical Decision Making 18(4): 418-428.

Tversky, A., D. Kahneman and R. Choice (1981) 'The Framing of Decisions', Science 211: 453-458.

Tversky, A. and D. Kahneman (1986) 'Rational Choice and the Framing of Decisions', The Journal of Business 59(4, Part 2: The Behavioral Foundations of Economic Theory): S251-S278.

Varey, C. and D. Kahneman (1992) 'Experiences Extended Across Time: Evaluation of Moments and Episodes', Journal of Behavioral Decision Making 5(3): 169-185.

Weinstein, M.C., G. Torrance and A. McGuire (2009) 'QALYs: The Basics', Value in health 12(s1): S5-S9.