The Effect of Group Decision-Making on Health Utility Biases

Anna James – 359576

Supervisor: Han Bleichrodt

22 October 2015

Erasmus University Rotterdam

M.Sc. Behavioral Economics 2015

Table of Contents

Abstract

This paper explores the effect of group decision-making on the systematic biases present in health utility measurement. The experiment allows for the comparison between individual and group health utilities measured using the Time Trade-off (TTO) and Standard Gamble (SG) methods, the latter involving risk and the former being riskless. The results indicate a significant difference between the group and individual health utilities measured using the TTO method, but not the SG method. Furthermore, there is a significant difference between the SG and TTO utilities calculated for the grouped participants, but not the individual subjects. However, when the SG utilities are adjusted for prospect theory, a significant difference is found between the TTO and SG elicited health utilities for the individual participants, but not for the grouped subjects, indicating a reduction in biases is present in the grouped participants for the TTO method, but not when making use of the SG method.

Introduction

Decision-making within healthcare is an intricate process involving several stakeholders with varying levels of expertise, from the patient themselves to a team of surgeons with different specializations. There is, however, one clearly overarching objective, which is to reduce the discomfort of the patient to a minimal level, especially in case of chronic conditions for which a cure is not possible. Although the patients' care is the number one priority, cost effectiveness must also be considered when assessing different health policies. In order to reach both goals from a more standardized point of view, it is first necessary to calculate and evaluate the utility of the specific health state in the patient's mind. Although this is a personal preference, there is a lot of information that a medical professional provides that will have a large effect on the evaluation of the health state. Furthermore, it is virtually impossible for an individual to make a decision concerning their health without an active discussion with their medical professional, or without gaining the opinion of family members or friends, especially when referring to chronic medical conditions. Thus it is illogical to measure health utilities on which medical decisions are assessed through isolated questionnaires, given that there are significant biases that affect individual decision-making. In order to gain a more realistic indication of the value assigned to a specific health state, a different approach should be taken that includes a collaborative method.

The challenge surrounding medical decision-making is the wide variety of cases, and the high variation even between cases with the same diagnosis. In the proposed study, the focus will lie primarily on the comparison between a decision made by an individual compared to one made in a pair, concerning a strictly defined health state to avoid further confusion. In order to define shared decision-making, several criteria must be met. These include the involvement of at least two participants, participation towards a shared decision by both, information sharing, and agreement on the final decision (Charles et al., 1997). Thus, the terms paired decision-making and group decision-making will be used interchangeably throughout this paper. The following literature review will outline the current methods used to elicit health utilities, the biases that arise from making use of these methods, and the strengthening of decision making through group cooperation.

Health Utilities

Firstly, it is important to understand what health utilities are, how they are measured and how they are used in medical decision-making. Health utilities are values customarily between 1.0 (healthy state) and 0.0 (death state) that indicate the strength of individuals' preferences for certain health related outcomes (Torrance, 1986). More specifically, health utility assessment is the process of assigning a quality weight to a specific health state, which then plays an important role in cost-effectiveness analysis (Doctor, Bleichrodt & Lin, 2010). One of the leading methods makes use of utility theory, and the preferences are thus considered utilities due to the fact that they are measured under uncertainty, whereas other methods determine values provided under certainty, eliminating the individual's risk attitude (Torrance, Furlong & Feeny, 2002).

The three most common direct measures of health utility include the standard gamble (SG), the time trade-off (TTO), and the visual analog scale (VAS). For the purpose of this research, the visual analog scale will not be covered in depth since it does not require that the subjects make a trade-off (choice) between different arguments, and thus is viewed by economists as inferior to the SG and TTO methods (Dolan, 2000). Furthermore, the SG and TTO methods are suitable for a variety of applications, including cost-utility analysis, quality adjusted life years (QALYs) and monitoring the health of a population (Torrance, Furlong & Feeny, 2002). The QALY model takes into account both the quality (health state utility) and quantity of life generated by a healthcare intervention, and is thus a measure of health effectiveness to be included in cost-effectiveness analysis (Weinstein, Torrance & McGuire, 2009).

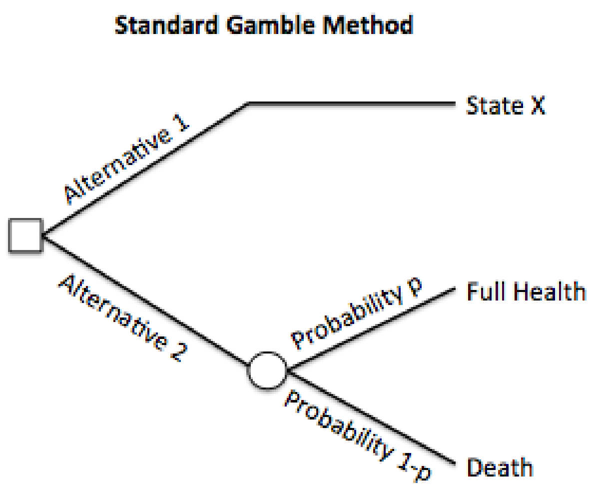
Given that health has a large influence on one's utility function, both methods are founded on the principle that it is possible to measure the welfare change associated with a change in health, if the compensating change in another factor in the utility function can be determined (Dolan, 2002). The methods can be seen as sharing a common theoretical background if it is assumed that improvements in health are a negative function of risk for the SG, and a positive function of longevity for the TTO method (Dolan, 2002).

Standard Gamble Method

There are many possible variations of the SG method, the essence being that a subject must make a choice between two alternatives. The first maintains a chronic health state (X) with certainty for T years, while the second offers the probability *p* of living in full health (FH) for T years and probability *1-p* of death immediately. By varying the value of *p* until the subject is indifferent between the two alternatives we uncover the health utility U(X) = *p* (Torrance, 1986). The probability *p* indicates the point of indifference between the certain chronic state (X) and said probability of living in full health. When this indifference is evaluated by expected utility theory the following is obtained:

U(T years in health state X) = *p* U(T years in FH) + (1-*p*) U(death)

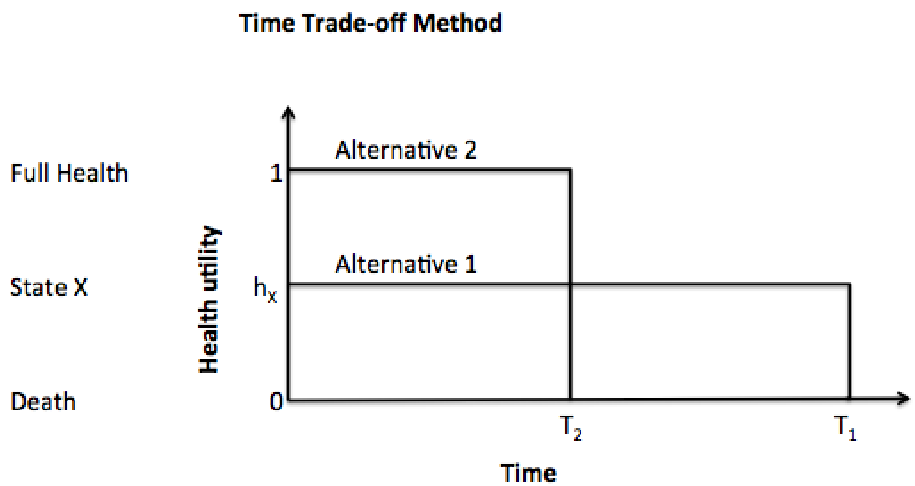
Thus given the scaling of U(T years in FH) = 1 and U(death) = 0, the resulting utility of the chronic health state is U(X) = *p*.



Time Trade-off Method

Alternatively, the TTO method involves a tradeoff between living with a chronic condition, and living for a shorter time at full health. As can be seen in the figure below, the subject must make a decision between state X (chronic health state preferred to death) for T_1 years, and full health (FH) for T_2 years ($T_2<T_1$). In this case T_2 is varied until the subject is indifferent between the alternatives. Finally, the specific health state utility is

computed by $U(X) = T_2/T_1$ given that the utility is linear in duration (Torrance, 1986). This implies that giving up years of life now rather than later will have an equal effect on an individual's utility, since each year of life yields the same utility (Dolan & Jones-Lee, 1997). This assumption will be discussed in depth in the following section.



The two methods outlined above are those most commonly used. However, they often provide different quality weights to the same given health state (Krabbe, Essink-Bot & Bonsel, 1997). A summary of the necessary notation for both methods can be found in Appendix A.1.

Health Utility Biases

Although the methods described above are standardized methods that have been used for many years, they are not immune to human biases. This is apparent from the differing results obtained from the different methods while measuring the utility of the same health state (Hornberger, Redelmeier & Petersen, 1992). Biases are the result of making use of heuristics in order to simplify tasks, leading to systematic error (Tversky & Kahneman, 1974). Essentially, these biases stem from violations of expected utility, and the three identified as leading to systematic error in either the SG or TTO methods include, probability weighting, loss aversion, and scale compatibility (Bleichrodt, 2002). Utility curvature was also identified as affecting the elicited health utilities using the TTO method, but stems from the violation of assumption of linear duration rather than a violation of expected utility.

For the purpose of this study, the focus will lie on the difference in health utilities obtained through the SG and TTO. The latter has consistently been proven to elicit lower individual quality weights than the SG method through a number of empirical studies (Torrance, 1976). In order to understand this difference in results, it is important to consider the systematic error effects mentioned above. A previous study by Bleichrodt determined the effect of each bias on the utilities elicited through the SG and TTO methods, as can be seen in the table below (2002). Each one will be covered individually in the following sections, addressing the reason for the direction of the bias.

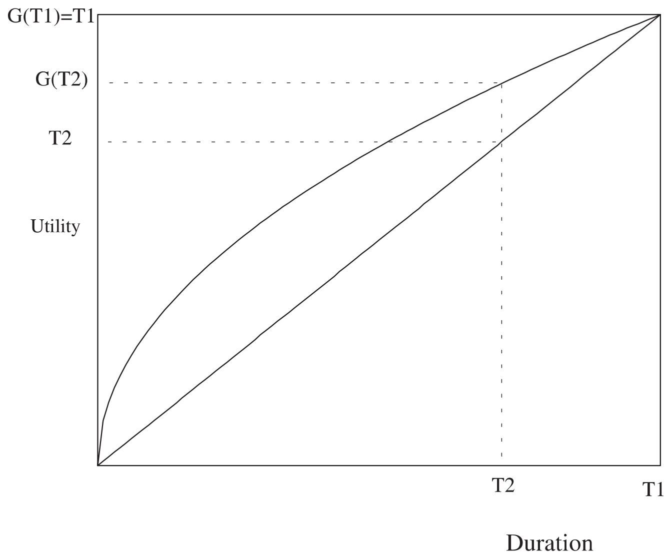
Effect	Bias in SG utility	Bias in TTO utility
Utility curvature	None	Downward
Probability weighting	Generally upward	None
Loss aversion	Upward	Upward
Scale compatibility	Ambiguous	Upward

[1]

[1] Overview of the biases in the SG and TTO utilities (Bleichrodt, 2002)

Utility Curvature

The first effect discussed is only relevant for the TTO method since there is an assumption of linear utility in duration, which is not the case in the standard gamble. This assumption of linear utility for duration leads to downward bias in TTO utilities, given a concave utility function over duration. This is due to the fact that for most people the marginal utilities for length of life vary from year to year, increasing more rapidly in the short term compared to long term (Stiggelbout et al., 1994). In layman's terms, people are more concerned with their short-term health than long-term health, and will act accordingly when making decisions concerning their well-being. Dolan and Jones-Lee indicate that a participant completing a TTO task will be more willing to sacrifice more years in the future rather than years of life now, thus having a positive rate of time preference and a concave utility function over duration, as seen below (1997). This downward bias is thus intensified in utility measurement concerning one's health in the short-term, leading to higher inaccuracy in the elicitation.



[2]

[2] Concave utility function indicates the downward biased TTO utilities (Bleichrodt, 2002)

Probability Weighting

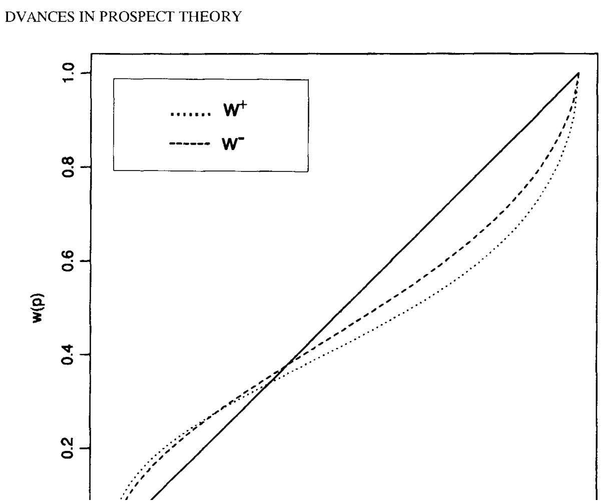
This bias, on the other hand, does not affect the TTO method since there are no probabilities involved. Therefore, the focus lies on the SG method and the probabilities this method relies on. Expected utility assumes that people evaluate probabilities in a linear fashion. However, this is not the case and there is a vast amount of empirical evidence indicating that people assign weights to different probabilities, and these are demonstrated by an inverse S-shaped probability weighting function as seen below (Tversky & Kahneman, 1992).

One non-expected utility theory that allows for the non-linear weighting of probabilities is rank-dependent utility (RDU) theory, and does so by creating a decision weight (*w*) of an outcome that is a function of the probability of the outcome along with its rank compared to the other outcomes (Bleichrodt, van Rijn & Johannesson, 1999). Thus under RDU, the SG utility of health state X becomes as follows:

U(T years in health state X) = $w(p)$ U(T years in FH) + $w(1-p)$ U(death)

Thus given the scaling of U(T years in FH) = 1 and U(death) = 0, the resulting utility of the chronic health state becomes U(X) = $w(p)$.

The point at which the probability weighting function switches from concave to convex occurs between .3 and .4, leading to an upward bias given that $w(p) < p$ and the probabilities are thus underweighted. Since the standard gamble method used in this study only presents options with percentages greater than .5, the focus lies on the upper half of the probability weighting function. The health utilities elicited through the SG method are thus overstated due to the underweighting of larger probabilities.



[3]

Loss Aversion

One of the more familiar deviations from expected utility introduced by Tversky and Kahneman is the evaluation of outcomes relative to a reference point with a higher sensitivity to losses than to gains known as loss aversion (1979). It is logical that a bias such as loss aversion would play a role in decisions concerning one's health and duration of life, given the high value assigned to one's own life. In the following sections the reference points used in both methods will be explored.

In the time trade-off method, the participant must indicate how many years in full health (T_2) will make him indifferent to a given number of years (T_1) in a chronic health state. Loss aversion implies that one will take T_1 as a reference point, and be less willing to give up years of life in order to gain health status, thus creating an upward bias in the TTO elicited health utilities. However, as Bleichrodt (2002) points out, this effect is reversed by framing the question differently. If an individual is asked to give up health to gain years on their life, this effect will be reversed and the TTO utilities will be biased

[3] Probability weighting function for gains (w^+) and losses (w^-) (Tversky & Kahneman, 1992)

downwards. In that situation, the reference point becomes the state of full health, and a subject will be less willing to give up what they already perceive to have.

In the standard gamble method, the certain chronic health state (X) is the reference point when it comes to evaluating the options presented. The choice presented to the subject is thus the decision between T years in a chronic health state versus the probability (p) of living the same number of years in full health, with probability (1-p) of immediate death. Given that the reference point taken will be the certain option (T years in a chronic health state), a higher probability (p) of living in full health will be necessary for a loss averse subject to renounce the certainty of the other option. This can be attributed to the fact that the loss of years of life has a greater impact than the gain in overall health, and thus a higher chance of full health is necessary to offset the loss of certainty. This behavior indicates a systematic upward bias in the health utility elicited through the SG method in case of loss aversion.

Scale Compatibility

The final effect addresses the violation of procedural invariance, which is the assumption that people have a well-defined utility function leading them to provide the same preference ordering through different elicitation techniques (Tversky, Sattath & Slovic, 1988). Delquiè (1993) proved that this assumption is systematically violated, triggering research to determine various explanations. Scale compatibility provides an explanation for the violation of procedural invariance by demonstrating that one's preferences are dependent on the response scale used and an individual attaches more weight to an attribute given its higher compatibility with the response scale used.

Given that the response scale used in the TTO method is in years, scale compatibility indicates that the subject will place more weight on the information provided concerning duration rather than health status. Therefore, with the focus laying on duration, the participant will be less willing to give up years for an improvement in health status, leading to an upward bias in the elicited health utilities.

Alternatively, the response scale used in the SG method is probability, shifting the subjects' weight to those provided. However, the direction of this bias is not clearly defined, since it is impossible to determine whether a participant will put more weight on the probability of full health (p), the probability of death ($1-p$) that are provided, or the certainty ($p = 1$) of the certain option. The direction of this bias on health utilities elicited using the SG method is ambiguous, since which probability the subject focuses on will lead to opposite effects on the health utilities (Bleichrodt, 2002).

Group decision-making

Group decisions play a large role in everyday life and it is common for people to consult with others, expert or not, in order to make a difficult decision. Therefore, it is logical that this would be especially present in the case of medical decision-making, given the high value people assign to their overall well-being and the complexity that such decisions involve. A large body of research can be found concerning the performance of groups compared to individuals, with varying results for different types of tasks.

Pagliari et. al. defined the group decision making process into 3 separate stages of orientation, evaluation and control (2001). The first involves the group members defining the problem at hand. They must then evaluate the options, which is where the most discussion takes place, and finally come to an agreement concerning the alternative to select. Within these three stages the group is susceptible to social influences from one another, which can lead to biases associated with group dynamics. It is not common for an individual to make decisions concerning their health without further discussion with a medical professional or friends and family, meaning that external players influence their decisions positively or negatively. For example, certain motivation losses are associated with group decision-making, commonly known as social loafing (Kerr & Tindale, 2004). This refers to the reduced effort of certain members, given the reduced risk of evaluation and opportunity to rely on others to complete the work. However, social loafing is moderated by high task involvement, thus less common in smaller groups with high engagement (Smith, Kerr, Markus & Stasson, 2001).

For the purpose of this research, we will explore whether group decision-making intensifies the biases found in individual decision-making. Kerr, Kramer and MacCoun investigated the bias they denote as *Sin of Omission*, which refers to the effect of framing in terms of gains or losses on one's decision (1996). Although mixed results were found, there was stronger evidence to the reduction of the bias through group collaboration. Furthermore, it was found that, on average, groups outperform individuals on various decision-making tasks (Kerr, MacCoun & Kramer, 1996).

Charness, Karni and Levin explored how moving away from the unrealistic laboratory environment, where decisions are made in isolation, affects the rationality of the decisions made (2007). To this end, they studied the hypothesis that decisions made by groups are more likely to be to be correct compared to individual decision makers. The results indicate that violations of stochastic dominance and Bayesian updating are reduced through social interaction, and decrease with the growing size of the group (Charness, Karni & Levin, 2007). Charness and Sutter also studied the evidence that groups have a higher chance of making rational decisions, following standard game-theoretic predictions, compared to individuals who are more affected by social considerations, cognitive limitations and biases (2012). This can be attributed to different factors, such as the reduced effect of bounded rationality and increased self-interest behavior found in groups.

Another aspect to consider in relation to the SG method is the comparative risk attitude of groups and individuals. The empirical studies comparing decision-making under risk for groups and individuals have thus far found mixed results. Rockenbach, Sadrieh & Mathauschek (2007) found that there is little difference in the rates of violation of expected utility between groups and individuals. There is also evidence supporting both claims that groups are less risk averse than individuals (Zhang & Casari, 2009), and more risk averse than individuals when high levels of risk were involved (Shupp & Williams, 2008). Furthermore, other studies found no differences between the risk attitudes of groups and individuals (Harrison et al., 2013).

There is currently no literature addressing group decision-making directly in relation to health utility measurement. Therefore, the combination of the research on systematic biases present in health utility measurement and the evidence supporting group decision-making, leads to the following experiment exploring the reduction of systematic biases in health utility measurement through group decision-making. The possibility that performing SG and TTO tasks in a group will result in the reduction of systematic biases found in individuals' preferences will be studied.

Methodology

The literature indicates an overall mixed opinion on the benefits of group decision-making due to the wide variety of situations examined, especially in decisions involving risk. However, the research addressing the reduction of biases through group decision-making and the nature of the collaboration necessary in this experiment leads me to the following hypothesis:

$H_{1:}$ The grouped subjects will demonstrate a reduced difference between the SG and TTO utilities compared to the individual subjects.

In order to measure the difference in bias through group decision-making, an experiment was conducted involving one group of individual participants and one group of pairs. These groups were given the same tasks, an SG and TTO task to determine their indifference points, and from that measure their health utilities for the given health state. According to Torrance (1986), this process requires 4 steps, namely:

1) Health state identification

The health state that was used for both the SG and TTO methods is the EQ-5D 22122, which describes a state of chronic moderate back pain. This health state was chosen since it is not an uncommon ailment and subjects can therefore relate to the symptoms. More importantly was the time in years (also known as gauge duration) that the subjects would be affected by this ailment, since this can affect their decision greatly. As found by Bleichrodt, Pinto, and Abellan-Perpiñan, the inconsistencies found in the TTO method due to loss aversion decrease with a higher gauge duration, and TTO exceeded SG utilities only when short gauge times were used (2003). Therefore, in this experiment 24 years was chosen as the gauge duration for both tasks, given that the subjects are students, and could find it challenging to realistically judge much longer gauge times.

2) Health state description

The description of the EQ-5D 22122 health state is and was presented as follows:

- Some problems walking about
- Some problems performing self-care activities
- No problems performing usual activities
- Moderate pain or discomfort

No further information was given concerning the health state, and the participants were instructed to imagine they were currently living in this health state. The description is intended to have a functional focus, rather than providing the clinical details of the health state, as described by a physician (Torrance, 1986). This specificity allows for the subjects to judge the health state without room for their own interpretation or confusion due to uncommon medical terms.

3) Subject selection and Incentives

The 64 subjects selected were students of Erasmus University Rotterdam, recruited in person at university or invited to participate via means of social media. All participants were given the option to qualify for the chance of winning a 20€ Bol.com gift card as an incentive to participate by providing their e-mail address after having completed the tasks. Those completing the two tasks in a group were recruited in person at the university, thus ensuring that the members of a group knew one another beforehand. This was done purposefully in order to enable a certain level of discussion that may not occur between strangers.

4) Use of utility measuring instruments

As discussed previously in detail, the instruments used are two choice lists presented in a table, one making use of the SG method and the other of the TTO method. The forms given to the subjects can be found in Appendix B. In the time trade-off method, the choices given involve durations that are not multiples of 5, since subjects are more likely to choose round numbers over non-round numbers (Bleichrodt, Pinto &

Abellan-Perpiñan, 2003). On the other hand, the SG method presents probabilities in round numbers at equal distances, since it can be harder to process percentages.

Procedure

After the students had agreed to participate, they were instructed on the choice list method used and the pairs were explicitly told that they should discuss the decision before filling out the form. Finally, they were left to complete the task on paper, which was then picked up once they had finished. Due to time limitations and lack of willing subjects present at the university during the summer holidays, the individual participants filled out the exact same form, but through an online questionnaire tool. They were incentivized through the same offer of a chance to win a 20€ Bol.com gift card by providing their e-mail address at the end of the two tasks. The forms shown were identical, whether administered online through Qualtrics or in paper form, thus the content did not affect their responses. Furthermore, Ross et al. developed a Paper Standard Gamble (PSG) questionnaire that would be simple to administer, and found no difference between the SG and PSG methods, concluding PSG to be a reliable substitute for the standard SG method (2003).

Statistical Methods

In order to test the hypothesis, $H_{1:}$ *The grouped subjects will demonstrate a reduced difference between the SG and TTO utilities compared to the individual subjects*, it was firstly necessary to calculate the individual health utilities from the switching points indicated by the subjects for each method.

From a quick glance at the table below summarizing the average health utilities indicated per method and group, the SG method appears to have a higher average compared to the TTO, and the individual subjects' average health utility appears higher. In the following section the statistical tests used will be discussed to determine the significance of these differences. Due to the unknown distribution of the populations, non-parametric tests were chosen to analyze the data.

Average HU	**Standard Gamble**	**Time Trade-off**
Individual	0.875	0.8479
Group	0.8318	0.5777

Between Subject Analysis

Firstly, two Mann-Whitney U tests were performed to determine whether there was a significant difference between methods for groups and individuals (Appendix C1.1).

There was a significant difference between the health utilities elicited using the TTO methods for groups and individual participants, at a 1% significance level with p = 0.0001. However, this was not present in the SG health utilities, where the resulting p-value was 0.4824. In fact, there was essentially no difference between the two groups for the SG method, indicating that the upward biases were not removed through executing the task in collaboration with another person. This result indicates that the upward biases in the TTO method caused by loss aversion and scale compatibility were likely reduced through discussing the choice with another person.

Within Subject Analysis

A Wilcoxon test was performed to assess whether there was a significant difference within groups, thus whether the individual participants' SG health utilities were significantly different from their TTO health utilities, and similarly for the grouped participants.

The results found in Appendix C1.2 indicate that the grouped participants demonstrated significantly different health utilities elicited using the two different methods at a 1% significance level, with p = 0.000. The individual participants, on the other hand, resulted in a p-value of 0.1570 and thus did not show a significant difference. These results are contrary to the hypothesis, given the expectation that the difference would be greater in the individual participants, whereas this indicates that the opposite is true. However, as expected from the available literature (Torrance, 1976), the SG utilities were higher than the TTO elicited for both groups and individuals.

In order to test the hypothesis more precisely, a Mann-Whitney U test was performed on *hu_diff*, indicating the difference between SG utilities and TTO utilities, with grouping variable *individual*, that specifies whether the participants completed the task alone (1) or in a pair (0).

The output indicates that there is in fact a statistically significant difference between SG and TTO for the pairs and individuals at a 1% significance level, with p = 0.000. Furthermore, from the resulting rank sums, it is apparent that the group difference is significantly higher than the individual difference, contrary to the prediction.

Adjusted SG Utility Analysis

Due to the findings in literature indicate that the SG method health utilities are consistently biased upwards, adjusted SG utilities were calculated in accordance with prospect theory. Kahneman and developed this alternative to the expected utility model for decision-making under risk in which probabilities are replaced by decision weights, given the systematic violation of expected utility axioms (1979). Bleichrodt, Pinto and Wakker explored the use of prospect theory to correct for the common violations of expected utility under risk and uncertainty (2001). In order to quantify these adjustments, two deviations from expected utility were assumed, namely the non-linear weighting of probabilities and the tendency to assign more weight to gains than losses, known as loss aversion. The resulting corrected probability equivalent utilities that were used to adjust the SG health utilities can be found in Appendix C.2, and the average adjusted SG utilities are presented below.

Average HU	**Standard Gamble**	**Time Trade-off**	**Adjusted SG**
Individual	0.875	0.8479	0.6627
Group	0.8318	0.5777	0.6195

The average adjusted SG utilities are clearly much lower than the non-adjusted values, and for the individuals the average is lower than that of the individual TTO utilities. This difference indicates that the upward biases found using the SG method were not reduced through group decision-making, as they were with the TTO utilities. The significance of the difference between the adjusted SG utilities is explored below for both individuals and grouped participants.

The interesting result obtained from adjusting the SG utilities according to prospect theory is the reversal of the significant differences previously found. Where before being adjusted the difference between the group SG and TTO was statistically significant, it now results in no significant difference ($p = 0.2484$). This indicates that the group collaboration reduced the TTO elicited health utilities in the same way as the adjustment for prospect theory did the SG utilities. The differences between individual health utilities were previously found to be non-significant, whereas the prospect theory adjusted SG health utilities are significantly difference from the TTO utilities at a 1% significance level, with $p = 0.0003$.

Results

Whether we consider the SG health utilities adjusted for prospect theory or not, there is not a significant difference between the values elicited for individuals or grouped participants. This indicates that the biases present in the SG elicitation procedure were not eliminated or reduced through group collaboration. This is also apparent when comparing the SG utilities to those adjusted for prospect theory, since there is a strong upward bias in the original values elicited, for both the individuals and grouped participants. The table below provides an overview of the statistical tests performed and those that were found to be significant at a 1% significance level.

Test	Sample 1	Sample 2	Significant at 1%
Ranksum	SGg	SGi	No
Ranksum	TTOg	TTOi	Yes
Wilcoxon signed-rank	SGg	TTOg	Yes
Wilcoxon signed-rank	SGi	TTOi	No
Wilcoxon signed-rank	TTOg	Corrected SGg	No
Wilcoxon signed-rank	TTOi	Corrected SGi	Yes

The fact that there is not a significant difference between the group TTO utilities and the adjusted SG values, along with the significant difference between the same values for the individual participants, indicates that the collaboration led to a reduction of biases in the TTO method. Since both methods are known to lead to utilities with an upward bias in individual participants (Torrance, 1976), the fact that the grouped TTO was much lower is in line with the expectations stated in the hypothesis, unlike the SG health utilities.

It is likely that the nature of the methods is responsible for the difference in the results. The SG method involves risk and the percentages represent this risk for each option. This can have an effect on the benefit of collaboration, since people can have very different risk attitudes. If one member is more risk averse, it will be much harder to convince them to take a more risky option regarding their own well-being, compared with the opposite situation, especially since groups are known to be more risk averse than individuals when confronted with high levels of risk (Shupp & Williams, 2008). Even if one subject is generally more risk seeking, it is simpler to convince someone to select an option that will protect his or her own well-being. Furthermore, previous studies have found little evidence that group decision making under risk leads to greater compliance with expected utility theory (Rockenbach, Sadrieh & Mathauschek, 2007). Bone et al. also demonstrated that groups and individuals had similar rates of violation of expected utility when making decisions under risk (1999). In the TTO method, on the other hand, there is not a risk factor but simply a decision based on certain outcomes. This situation could enable more discussion, leading to the results seen above of reduced biases in the group setting.

Conclusion

The field of medical decision-making and group decision-making are developed fields that each provide a large amount of information, yet the link between the two fields has purely addressed the psychological effects of groups acting in a medical setting. With this experiment, I explored the effect of group decision-making specifically on the measurement of health utilities, in order to provide further insight into the reduction of specific biases through collaboration. The initial results were ambiguous, indicating a significant reduction through the riskless TTO method, but a lack of a significant effect through the SG method involving risk.

The contrast between the elicited SG utilities before and after the adjustment according to prospect theory indicates how greatly the original values are biased upwards. It also allows for a more reliable comparison between the SG and TTO elicited utilities, although the latter are also highly inflated in the individual subjects. It is apparent that the biases found using the TTO method are highly reduced through a group decision-making process, supporting the initial hypothesis.

From comparing the two elicitation methods, the significant difference between the adjusted individual SG utilities and the individual TTO values represents the highly biased nature of the latter. Furthermore, the fact that these health utilities are not significantly different for the grouped participants indicates that the grouped TTO method does in fact reduce the biases present, while the (unadjusted) SG method does not.

A possible explanation concerning the group TTO method reducing biases present and the group SG method not is that the former is much simpler for participants to understand. The probabilities involved in the SG elicitation method could create more confusion amongst participants. Furthermore, in a group it is likely that different members have different risk attitudes, and it is generally less challenging to convince someone to be more loss averse rather than less, leading to the same biases present in the individual participants' health utilities. It is also possible for the participants to have differing time preferences that could affect their responses in the TTO method. However

Sozou and Seymour provide evidence that a U-shaped relationship exists between age and time preference (2003), thus leading to believe that this would not have a large effect on the results given that the participants fall within the same age range.

Limitations

One clear limitation of this study was the setting in which the participants were asked to complete the task. Due to limited resources, it was not possible to measure the health utilities in a lab setting, thus allowing for the distraction of the participants. Furthermore, since the groups were approached while already sitting together, it is very likely that the group members knew each other already. This could lead to a higher level of trust between them and therefore influence their decisions accordingly.

Another factor that could have influenced the results is the fact that the individual participants completed the task online, meaning the setting was most likely different than for the groups as university. Although the task was identical and there is evidence supporting the equality of the results obtained through online and paper methods, the vocal instructions provided to the grouped participants could have led to a difference in their overall understanding of the task (Ross et al., 2003).

It would also have been beneficial to sample a larger number of participants in order to gain a more representative result and increase the power of the tests. More importantly, the number of subjects in one group performing the task was always two. Increasing this number to 3 or more could have influenced the group dynamics and affected the results as well, since violations of stochastic dominance and Bayesian updating are reduced through social interaction, and decrease with the growing size of the group (Charness, Karni & Levin, 2007). However, a larger group presents more biases associated with group dynamics, which could lead to various effects on the results. Furthermore, it would have been interesting to have the participants complete the tasks individually as well as in a group to control for between subject confounding effects.

Suggestions for Future Research

In previous studies by Charness, Karni and Levin (2007), the biases associated with individual decision-making were found to decrease with group size. Therefore, it would be highly beneficial to conduct this experiment in a controlled environment with groups formed by three or more participants.

Given the results from the comparison of SG and TTO health utilities, a study about the ease of persuasion to taking a more risk averse attitude would be informative to determine what lead to the difference in group decision-making. It is logical that people are more easily convinced in a social setting to take a more risk averse stance, given that the decisions affects others, not only their own well-being (Zhang & Casari, 2009). An individual will not feel entitled to push others to select a more risky option since the outcome will have a more widespread effect in case of a negative outcome.

Another adjustment that could improve the validity of the study would be to create a medical setting in which the subjects would complete the task. Given that health utilities are being elicited, it would be interesting to explore the effect of the presence of a medical professional on the resulting health utilities. Lastly, the results of this study address only one chronic health state with consequences that are not extremely dire for the patient. It is recommended to explore the effect of group decision-making on health utility elicitation in different settings and with a range of health states to gain a more reliable insight into its effect on the resulting utilities.

References

Bleichrodt, H. (2002). A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Econ.*, *11*(5), 447-456. doi:10.1002/hec.688

Bleichrodt, H., Pinto, J., & Maria Abellan-Perpiñan, J. (2003). A consistency test of the time trade-off. *Journal Of Health Economics*, *22*(6), 1037-1052.

Bleichrodt, H., Pinto, J., & Wakker, P. (2001). Making Descriptive Use of Prospect Theory to Improve the Prescriptive Use of Expected Utility. *Management Science*, *47*(11), 1498-1514.

Bleichrodt, H., van Rijn, J., & Johannesson, M. (1999). Probability Weighting and Utility Curvature in QALY-Based Decision Making. *Journal Of Mathematical Psychology*, *43*(2), 238-260.

Bone, J., Hey, J., & Suckling, J. (1999). Are groups more (or less) consistent than individuals?. *Journal Of Risk Uncertainty*, *8*, 63-81.

Charles, C., Gafni, A. and Whelan, T. (1997). Shared decision-making in the medical encounter: What does it mean? (or it takes at least two to tango). *Social Science & Medicine*, 44(5), pp.681-692.

Charness, G., Karni, E., & Levin, D. (2007). Individual and group decision making under risk: An experimental study of Bayesian updating and violations of first-order stochastic dominance. *Journal Of Risk And Uncertainty*, *35*(2), 129-148.

Charness, G., & Sutter, M. (2012). Groups Make Better Self-Interested Decisions. *Journal Of Economic Perspectives*, *26*(3), 157-176.

Delquié, P. (1993). Inconsistent Trade-Offs Between Attributes: New Evidence in Preference Assessment Biases. *Management Science*, *39*(11), 1382-1395.

Doctor, J., Bleichrodt, H., & Lin, H. (2010). Health Utility Bias: A Systematic Review and Meta-Analytic Evaluation. *Medical Decision Making*, *30*(1), 58-67.

Dolan, P. The measurement of health-related quality of life for use in resource allocation decisions in health care. In Handbook of Health Economics, vol. 1B, Culyer AJ, Newhouse JP. (eds) Elsevier Science: Amsterdam, 2000; 1723–1760.

Dolan, P., & Jones-Lee, M. (1997). The time trade-off: A note on the effect of lifetime reallocation of consumption and discounting. *Journal Of Health Economics*, *16*(6), 731-739.

Harrison, G., Lau, M., Rutström, E., & Tarazona-Gómez, M. (2013). Preferences Over Social Risk. *Oxf. Econ. Pap. 65*, 25–46.

Hornberger, J., Redelmeier, D., & Petersen, J. (1992). Variability among methods to assess patients' well-being and consequent effect on a cost-effectiveness analysis. *Journal Of Clinical Epidemiology*, *45*(5), 505-512.

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

Kerr, N., MacCoun, R., & Kramer, G. (1996). Bias in judgment: Comparing individuals and groups. *Psychological Review*, *103*(4), 687-719.

Kerr, N., & Tindale, R. (2004). Group Performance and Decision Making. *Annual Reviews Psychology*, (55), 623-655.

Krabbe, P., Essink-Bot, M., & Bonsel, G. (1997). The comparability and reliability of five health-state valuation methods. *Social Science & Medicine*, *45*(11), 1641-1652.

Pagliari C., Grimshaw J., Eccles M. (2001). The potential influence of small group processes on guideline development. *J Eval Clin Pract*. 7, 165–73.

Rockenbach, B., Sadrieh, A., & Mathauschek, B. (2007). Teams take the better risks. *Journal Of Economic Behavior & Organization*, *63*(3), 412-422.

Ross, P., Littenberg, B., Fearn, P., Scardino, P., Karakiewicz, P., & Kattan, M. (2003). PAPER STANDARD GAMBLE. *International Journal Of Technology Assessment In Health Care*, *19*(01).

Sozou, P. D., & Seymour, R. M. (2003). Augmented discounting: Interaction between ageing and time-preference behaviour. *Proceedings of the Royal Society of London Series B*, *270*, 1047–1053.

Shupp, R., & Williams, A. (2008). Risk preference differentials of small groups and individuals. *The Economic Journal*, *118*(525), 258-283.

Smith, B., Kerr, N., Markus, M., & Stasson, M. (2001). Individual differences in social loafing: Need for cognition as a motivator in collective performance. *Group Dynamics: Theory, Research, And Practice*, *5*(2), 150-158.

Stiggelbout, A., Kiebert, G., Kievit, J., Leer, J., Stoter, G., & De Haes, J. (1994). Utility Assessment in Cancer Patients: Adjustment of Time Tradeoff Scores for the Utility of Life Years and Comparison with Standard Gamble Scores. *Medical Decision Making*, *14*(1), 82-90.

Torrance, G., Furlong, W., & Feeny, D. (2002). Health utility estimation. Expert Review Of Pharmacoeconomics & Outcomes Research, 2(2), 99-108.

Torrance, G. (1976). Social preferences for health states: An empirical evaluation of three measurement techniques. *Socio-Economic Planning Sciences*, *10*(3), 129-136.

Torrance, G. (1986). Measurement of health state utilities for economic appraisal. *Journal Of Health Economics*, *5*(1), 1-30.

Tversky, A., Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal Of Risk And Uncertainty*, *5*(4), 297-323.

Tversky, A., Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124-1131.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, *95*(3), 371-384.

Weinstein, M., Torrance, G., & McGuire, A. (2009). QALYs: The Basics. *Value In Health*, *12*, S5-S9.

Zhang, J., & Casari, M. (2009). How Groups Reach Agreement In Risky Choices: An Experiment. *Economic Inquiry*, *50*(2), 502-515.

Appendix

Appendix A

A1 Notation

X	Chronic health state preferred to death
FH	Full health state
T_1	Time duration spent in state X
T_2	Time duration spent in FH
p	Probability of being in FH
$1-p$	Probability of death

Appendix B

B1 Health Utility Elicitation Tasks

Imagine that you are living with the following chronic back pain condition

Some problems walking about

Some problems performing self care activities

No problems performing usual activities

Moderate pain or discomfort

Task 1

Suppose that you have 24 more years to live with back pain.

In the following task, you must consider the following alternatives:

1) Living for 24 years with back pain, followed by death
2) Living for Y years in full health, followed by death.

Please indicate for each step the alternative you would prefer.

	Current Situation 1	Possible Situation 2	Decision	
Step	Years with back pain	Years in full health (Y)	Remain in 1	Switch to 2
1	24	1		
2	24	6		
3	24	11		
4	24	14		
5	24	19		
6	24	24		

Task 2

Suppose that you have 24 more years to live with back pain.

In the following task, you must consider the following alternatives:

1) Living for 24 years with back pain, followed by death
2) A risky treatment that gives a probability P of living for 24 years in full health and a probability (1-P) of immediate death.

Please indicate for each step which alternative you would prefer.

	Current Situation 1	Possible Situation 2		Decision	
Step	Years with back pain	Years in full health	Probability of full health (death)	Remain in 1	Switch to 2
1	24	24	50% (50%)		
2	24	24	60% (40%)		
3	24	24	70% (30%)		
4	24	24	80% (20%)		
5	24	24	90% (10%)		
6	24	24	100% (0%)		

Age __________

Gender M/F

E-mail address:

Appendix C

C1 Stata Output

C 1.1 Between Subject Analysis

```
. ranksum hu_SG, by (individual)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

  individual |      obs    rank sum    expected
-------------+---------------------------------
           0 |       22         446         473
           1 |       20         457         430
-------------+---------------------------------
    combined |       42         903         903

unadjusted variance     1576.67
adjustment for ties      -99.52
                     ----------
adjusted variance       1477.14

Ho: hu_SG(indivi~l==0) = hu_SG(indivi~l==1)
             z =  -0.703
    Prob > |z| =   0.4824
```

```
. ranksum hu_TTO, by (individual)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

  individual |      obs    rank sum    expected
-------------+---------------------------------
           0 |       22       321.5         473
           1 |       20       581.5         430
-------------+---------------------------------
    combined |       42         903         903

unadjusted variance     1576.67
adjustment for ties     -116.39
                     ----------
adjusted variance       1460.28

Ho: hu_TTO(indivi~l==0) = hu_TTO(indivi~l==1)
             z =  -3.965
    Prob > |z| =   0.0001
```

C 1.2 Within Subject Analysis

```
.  bys individual: signrank hu_SG = hu_TTO

----------------------------------------------------------
-> individual = 0

Wilcoxon signed-rank test

        sign |      obs   sum ranks    expected
-------------+---------------------------------
    positive |       22         253       126.5
    negative |        0           0       126.5
        zero |        0           0           0
-------------+---------------------------------
         all |       22         253         253

unadjusted variance       948.75
adjustment for ties        -3.00
adjustment for zeros        0.00
                      ----------
adjusted variance         945.75

Ho: hu_SG = hu_TTO
             z =   4.113
    Prob > |z| =   0.0000
```

```
-> individual = 1

Wilcoxon signed-rank test

        sign |      obs   sum ranks    expected
-------------+---------------------------------
    positive |       10         135        97.5
    negative |        5          60        97.5
        zero |        5          15          15
-------------+---------------------------------
         all |       20         210         210

unadjusted variance       717.50
adjustment for ties        -1.50
adjustment for zeros      -13.75
                      ----------
adjusted variance         702.25

Ho: hu_SG = hu_TTO
             z =   1.415
    Prob > |z| =   0.1570
```

```
. ranksum hu_diff, by (individual)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

  individual |      obs    rank sum    expected
-------------+---------------------------------
           0 |       22       654.5         473
           1 |       20       248.5         430
-------------+---------------------------------
    combined |       42         903         903

unadjusted variance     1576.67
adjustment for ties      -12.39
                     ----------
adjusted variance       1564.27

Ho: hu_diff(indivi~l==0) = hu_diff(indivi~l==1)
             z =   4.589
    Prob > |z| =   0.0000
```

C 1.3 Adjusted SG Utility Analysis

```
. bys individual: signrank hu_SG_adj = hu_TTO

--------------------------------------------------
-> individual = 0

Wilcoxon signed-rank test

        sign |      obs   sum ranks    expected
-------------+---------------------------------
    positive |       15         162       126.5
    negative |        7          91       126.5
        zero |        0           0           0
-------------+---------------------------------
         all |       22         253         253

unadjusted variance       948.75
adjustment for ties        -3.00
adjustment for zeros        0.00
                      ----------
adjusted variance         945.75

Ho: hu_SG_adj = hu_TTO
             z =   1.154
    Prob > |z| =   0.2484
```

```
-> individual = 1

Wilcoxon signed-rank test

        sign |      obs   sum ranks    expected
-------------+---------------------------------
    positive |        2           9         105
    negative |       18         201         105
        zero |        0           0           0
-------------+---------------------------------
         all |       20         210         210

unadjusted variance       717.50
adjustment for ties        -4.00
adjustment for zeros        0.00
                      ----------
adjusted variance         713.50

Ho: hu_SG_adj = hu_TTO
             z =  -3.594
    Prob > |z| =   0.0003
```

C2 Prospect Theory Corrected Probability Equivalent Utilities

Table 1 Corrected PE Utilities (By Equation 9) as Function of p for $p = 0.00, \ldots, 0.99$; e.g., the Corrected PE Utility for $p = 0.15$ Is 0.123

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.000	0.025	0.038	0.048	0.057	0.064	0.072	0.078	0.085	0.091
0.1	0.097	0.102	0.108	0.113	0.118	0.123	0.128	0.133	0.138	0.143
0.2	0.148	0.152	0.157	0.162	0.166	0.171	0.176	0.180	0.185	0.189
0.3	0.194	0.199	0.203	0.208	0.213	0.217	0.222	0.227	0.231	0.236
0.4	0.241	0.246	0.251	0.256	0.261	0.266	0.271	0.276	0.281	0.286
0.5	0.292	0.297	0.303	0.308	0.314	0.320	0.325	0.331	0.337	0.343
0.6	0.350	0.356	0.363	0.369	0.376	0.383	0.390	0.397	0.405	0.412
0.7	0.420	0.428	0.436	0.445	0.454	0.463	0.472	0.481	0.491	0.502
0.8	0.512	0.523	0.535	0.547	0.560	0.573	0.587	0.601	0.617	0.633
0.9	0.650	0.669	0.689	0.710	0.734	0.760	0.789	0.822	0.861	0.911

[4]

[4] Bleichrodt, Pinto & Wakker, 2001