UNIVERSITY OF OSLO
Institute of Health and Society

Anchoring in the Lead-Time Time Trade-Off:

Does the "Starting-point" Influence Preference Elicitation?

Master Thesis

Carl Haakon Samuelsen

August 20, 2011

Preface

Summary

Background: The standard TTO's different treatment of states worse than death (SWD) and states better than death (SBD) is by many viewed as a *prima-facie* cause for concern related to potential bias caused by using two different elicitation regimens. In response to this concern, Robinson and Spencer (2006) described an approach that made it possible to value SWD and SBD in the same exercise by adding a "lead-time" to the valuation tasks. There is a considerable literature suggesting that theoretically irrelevant factors influence preference elicitation trough heuristic processes, and that these processes may bias the resulting TTO values. If the lead-time extension serve as a heuristic anchor, the *a priori* choice of the lead-times initial position might have consequences on the respondents elicitation process, and if pronounced, cause of bias and error. Assessing the existence, the magnitude or the direction of the potential influence might provide researchers with valuable information for future LT-TTO studies. **Purpose:** To explore potential influence caused by the added "lead-time" component in the elicited LT-TTO values. **Methods:** A LT-TTO survey with manipulated "lead-time" starting points were run on a sample of the Norwegian population. The manipulations added starting points ranging from 0 to 10 years above the starting point of the comparator health state. The respondents received eight EQ-5D descriptions of hypothetical health states and valued them using the manipulated LT-TTO survey. The Kruskal-Wallis one-way analysis of variance by ranks was applied, and a robust regression with Huber and Tukey bi weights was fitted to detect potential influence attributable to the starting points. **Results:** A significant between group variation in the elicited results were detected by the statistical testing, and the robust regression revealed a positive linear relationship between the TTO-values and the lead-time starting points with a coefficient of 0.039. **Conclusion:** The findings suggest that the LT-TTO values are influenced by the starting point of the "lead-time" and that the method is sensitive to heuristic processes mediated by the "lead-times" initial positioning. The observed variation raises doubts w ether the LT-TTO leads us closer to an unbiased preference elicitation tool, or if it only trades one type of heuristic effect off for another. Since it is difficult to perceive that the dependency is a result of an *a priori* property of peoples' perceptions of health and time, anchoring should be considered when designing and interpreting LT-TTO studies.

Acknowledgements

Working on a master thesis is seldom a one person project. Even though the writing, the formulation of the ideas and the development of the logical structures inevitably is linked to the aspiring scholar's independent work and personal development; the process affects and involve those in the close vicinity to the student. My supervisor Ivar S. Kristansen have trough the prolonged process provided excellent advice, and exciting opportunities for hands-on-working-experience trough opening the doors for participation in a project for the Norwegian Research Council. Liv Ariane Augestad and Kim Rand Hendriksen, the designers of the study used in the thesis, deserves special thanks, not only did they lend me data from one of their most exiting studies, they also provided sound advice and tutoring when needed. I also wish to thank family and friends for support and understanding during the extended writing process.

Contents

List of Figures

List of Tables

1 Introduction

How much are you willing pay for a chocolate bar? —Your answer will probably depend on your preferences for chocolate, the time spent since your last meal or your *social security number*. The latter is a bold statement, but nevertheless true if we are to believe one of Dan Arielys' auction experiments where he found a correlation between participating executives social security numbers and their willingness to pay for a chocolate bar (Ariely et al., 2004). Even-though the correlation probably won't hold true outside the experiment, it leads to an important question: —If preferences for simple well-known consumption goods are sensitive to factors irrelevant to the outcome, what then about questions of a more complex and abstract nature? Rather, what if the question was: —How many years in perfect health are you willing to trade off in order to be indifferent between spending life α, in perfect health, and spending life β, in reduced health? A likely reply is that questions concerning morbidity and mortality generally are more difficult to answer than questions relating to everyday topics, and that they probably are more sensitive to irrelevant factors than questions of a more simple nature. Needless to say, a questions accuracy for capturing a subjects unbiased preferences are integral if the answers are used to inform decisions-makers when they are deciding on issues with consequences for peoples health and well-being.

The *Time Trade-Off* (TTO) (Torrance et al., 1972) is a technique devised to implicitly derive a subjects preferences for health outcomes based on their responses to decision situations (Torrance et al., 1982). One of the most significant uses of the TTO is to generate values for the generic health states that are used for outcome measurement in economic evaluations. Several nations including the Netherlands, UK, US, Japan and Denmark have used the TTO-technique to generate their EQ-5D tariffs (Devlin et al., 2011). The tariffs are then used as inter-diagnostic indicators of disease-status that inform policy-makers' decisions relating to distribution and prioritizing in health care. In other words—the method try to inform *"big"* decisions by summing up the results from numerous *"smaller"* decisions made by a representative sample of the population. For this to work, the *"smaller"* decisions must (a) be in line with the *"big"* decision, and (b) give a true representation of the respondents values. The former holds if the *"big"* and the *"small"* decisions are equivalent. The latter dwells on the assumption of procedural invariance; which assumes that the participants in a elicitation survey are insensitive to the utility search methods.

Even though the thought that superfluous variables influence decision-making seems to be well established in the literature, there have been few attempts to assess the TTO-family's susceptibility to them. This seems peculiar since the TTO-values, as derivatives of hypothetical decision problems easily could be imagined to be influenced by the same heuristics as every other decision. If this is the case, knowledge about the size and direction of these effects will provide important information for the further usage of the TTO-instruments. We will use the relatively new TTO-protocol *Lead Time Time Trade-Off* (LT-TTO) (Robinson and Spencer, 2006) as the experimental platform in this survey. Our purpose is to explore potential heuristic consequences of the added "lead-time" on the respondents elicitation process by providing the respondents with LT-TTO valuation tasks were the starting points have been manipulated.

2 Theory and Related Literature

2.1 The time trade-off

The TTO was developed in the early seventies to accommodate the need for an instrument that could yield a *"simple and easy-to-administer"* way to assign weights to Quality Adjusted Life Years (QALYs) for individuals preferences for changes in health status (Drummond et al., 2005). A QALY is a composite measure that merges the two dimensions quality-of-life (q) and quantity of life (T) into a single metric (Zeckhauser and Shepard, 1976). The concept assigns a health related QALY-weight that corresponds to the current health state experienced in each point of time lived by the individual. The utility (φ) an individual experiences in health state (i) with quality q during time T is then defined by the surface of the plane that emerges between q_i and T as time progress. It then follows that the health related utility experienced during a lifetime is determined by the integral

$$\varphi_i = \int_0^T (q_i(t))dt \tag{1}$$

Since the flow of time T can be assumed to be a constant in all feasible applications of utility measurement, the height of q_i determines the amount of health-attributed satisfaction an individual experience for a given duration of life-time. According to the formal requirements of utility theory, the weights must be elicited by using risk based choice, e.g. the standard gamble (SG) (Von Neumann and Morgenstern, 1944). There are however research that indicates that the TTO can be considered as a feasible elicitator for peoples health preferences (Dolan et al., 1996). The requirement is therefore treated more as a formality rather than a rule in the practical usage of the TTO-protocol.

In a TTO-survey, respondents are asked to indicate the amount of time (t_α) in full health (h_α) followed by death (Ω), they are willing to accept in order to be indifferent between spending t_α in perfect health, and spending the time (t_β) in reduced health (h_β) followed by death Ω. Keeping t_β fixed the respondents are allowed to vary t_α until they are indifferent between the two health states. (Δt_α) then denotes the subjects willingness to trade life-time for life-quality and vice versa. The preference score, or the QALY-weight attributed to the respondents' point of preferential indifference h_β is then found by solving the equation[1]

$$t_\alpha - \Delta t_\alpha = t_\beta U(h_\beta) \tag{2}$$

[1] Note that the mathematical notations used in this paper differ from the traditional treatment of the TTO. The motivation is that the inclusion of Δt_i (willingness to trade) provides a more intuitive presentation of the formula since notation now also includes the subjects willingness to forgo lifetime.

and with some elementary algebra, the ratio

$$U(h_\beta) = \frac{t_\alpha - \Delta t_\alpha}{t_\beta} \tag{3}$$

defines the preference score for health state h_β.

Figure 1: Conventinal TTO: States better than death

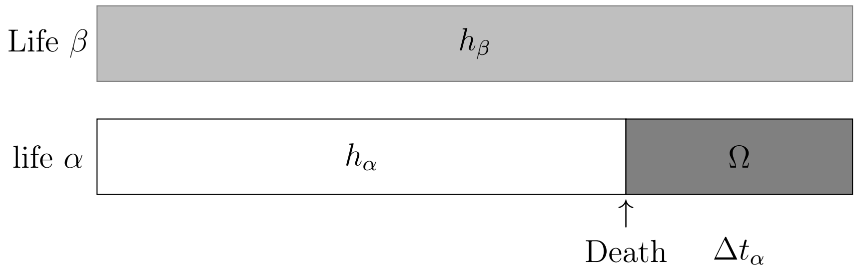
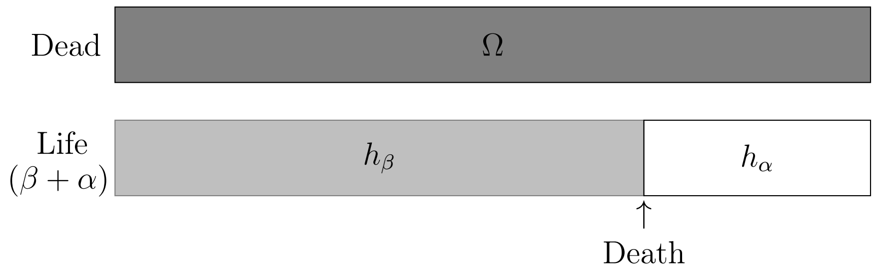


Figure 2: Conventinal TTO: States worse than death



If a respondent indicate he or she perceives a health state as worse than dead (SWD), an interviewer may provide the respondent with a different valuation task (TTO-SWD) that are able to measure negative valuations by asking respondents to choose between immediate death and spending t_β years in h_β followed by the time t_α in full health h_α, and keeping the total duration (T) of the two health states fixed. By the same logic as in (3 the utility of the SWD can be found by solving the equation)

$$U(h_\beta)t_\beta + \Delta t_\beta + t_\alpha + \Delta t_\alpha = 0 \tag{4}$$

Since T is a fixed integer we know that

$$T = (t_\alpha + t_\beta) \implies \begin{cases} t_\alpha = T - t_\beta \\ t_\beta = T - t_\alpha \\ \Delta t_\alpha = -\Delta t_\beta \implies \sum_{i=\alpha,\beta}^{t} \Delta t_i = 0 \end{cases} \tag{5}$$

Giving us the equation for SWD

$$U(h_\beta) = -\frac{T - t_\beta}{T - t_\alpha} = -\frac{t_\alpha}{t_\beta} \tag{6}$$

The procedural difference between the SWD and the SBD valuation tasks raises several concerns about the validity of the TTO-technique (Tilling et al., 2010). Since the valuation procedures of SBD and SWD are fundamentally different the aggregation of the values collected from SBD and SWD might be inconsistent. The different structures of the two valuation tasks might cause of a gap in the TTO-values often observed in health states with values close to zero (Stalmeier et al., 2005). Differences in the TTO values for SBD are obtained by varying t_α and holding t_β fixed. While on the other hand, the procedure for SWD involves simultaneously changing both the numerator and the denominator which makes the SWD values more sensitive to changes in small values as compared to the SBD. The TTO-values potential range of movement oscillates between one and negative infinity, this creates an asymmetry that introduce difficulties in interpreting the TTO-values. Devlin et al. (2011) points out that—76% of the states valued in the MVH study had negative mean values in the raw data before being transformed. Stalmeier et al. (2005) argues that once being transformed the SWD cannot be interpreted as utility scores. This could mean that only 24% of the material in the MVH study satisfies the already "informal" theoretical foundations that underlie QALYs elicited with the standard TTO. The "new" LT-TTO is an attempt to solve these issues by adding a *lead-time* that unifies the SWD and the SBD valuation tasks into one single operation.

2.2 The lead time trade-off

Figure 3: LT-TTO: States better than death

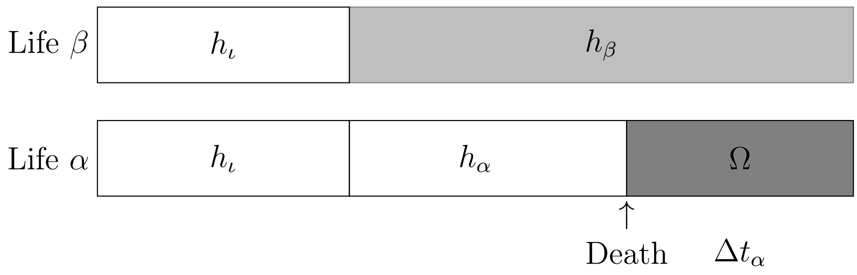


Figure 4: LT-TTO: States worse than death

The lead-time component (t_σ) attaches a pre-disease time-line specified as perfect health to the front of the traditional TTO. The added time increases the space available for varying t_α and enable respondents to iterate trough the disease onset when they perceive a health state as WTD $U(h_\beta) < 0$, without using two different TTO valuation tasks.

$$t_\sigma + t_\alpha - \Delta t_\alpha = t_\sigma + U(h_\beta)t_\beta \tag{7}$$

With some algebra we get the LT-TTO ratio

$$U(h_\beta) = \frac{t_\alpha - \Delta t_\alpha}{t_\beta} \tag{8}$$

The mathematical formulations (eqn. 3 and eqn. 8) suggests that the LT-TTO and TTO for SBD are the same by definition, the reason is surprisingly logical; unless we are dealing with peri-natal conditions there will always be a 'lead-time' in front of the t_i in real life. The problem have been that, until now, this opportunity have been unavailable. The equality holds as long as the lead time t_σ is the same for the scenarios h_α and h_β.

2.3 Decisions and heuristics

Everyday interactions with the surrounding environment provide humans with a nearly infinite amount of information (Dry et al., 2006). Despite the computational intractability of dealing with infinity, individuals seem to make their daily choices without an unreasonable amount of cognitive effort. The paradox is often illustrated by the traveling salesman problem (TSP) were a salesperson who wants to find the shortest route between a set of cities encounter an optimization problem. Common sense suggests that the problem is easily solved by just measuring the distances between the cities and opt for the shortest route—especially if we are dealing with say "only" fifteen cities. However, when we calculate that a salesperson planning to visit fifteen cities must choose between 43 billion possible routes, only one of them being optimal, the elusive character of the puzzle becomes evident. The reason is that the TSP is in the NP-complete (Non-deterministic Polynomial) complexity class, meaning that the time (t) needed to verify a solution increase with the polynomial of instances (n) such that $t = (1-n)!$ (Papadimitriou, 2003). The computational capacity needed to calculate the optimal (shortest) route between the 43 billion possibilities will most likely exceed the cognitive capacity of most human beings—it is therefore slightly peculiar that even small children are able to solve the problem within a satisfactory level of accuracy. In fact, Van Rooij et al. (2006) observed that a sample of 7-year old children were able to find solutions that on average were 7% above the optimal when solving sets containing 5-15 nodes. Compared with simple computer run optimization strategies such as "the nearest neighbor" or the "elastic net", the former having an average deviation of 25% from the optimal, and the latter averaging with an inefficiency of 8.5% (MacGregor and Chu, 2011) the children's performance must be considered as encouraging. Adult performance is slightly better, in small sets containing 6-10 nodes, adults on average find close to optimal solutions (Graham et al., 2000). In sets containing 10-25 nodes, the solutions are approximately 5% above optimal (Vickers et al., 2001). In a large set containing 120 nodes Dry et al. (2006) observed that the average performance were 11% above the optimal. They also observed that the human solution time increased linearly with number of cities, which is surprising given the NP-hardness of the TSP-problem.

By turning the TSP upside down and challenging the respondents to find the longest instead of the shortest route, (Chronicle et al., 2006) observed that the respondents' performance was substantially reduced. Endowed with a 10-node TSP, none of the respondents in the experiment were able to find the longest route. This was surprising since 31 of the 100 respondents earlier were able to find the shortest route on the same map. The average time used to solve the inverted TSP was 225% higher than the standard TSP (12.76 min vs. 5.67 min). The computer on the other hand, solved the two problems with equal ease and without any differences in accuracy using the same simple optimization algorithms, which indicate that the formal difficulty level of the two tasks were equal (Ibid).

One explanation to the dissonance of solution-quality between the similar problems may be attributed to the architecture of the cognitive system. Developed trough evolutionary processes the cognition is thought to resemble a "heterogeneous network of functionally specialized computational devices" designed to "solve specific tasks" rather than functioning as a "general purpose problem-solver" (Cosmides and Tooby, 1994). The specific task specialization provides a decision maker with domain specific tools that may give an edge over general-purpose strategies that are constrained by the need to apply the same problem-solving techniques for every encountered problem (Cosmides and Tooby, 1994). The ability to solve TSP-problems efficiently are thought to be an essential skill for mobile organisms relying on foraging or hunting as means to harvest nutrients necessary to sustain life. Furthermore, it is reasonable to believe that evolution has favored entities capable of solving TSP-problems fast and frugally. Todd (2001) provides an example explaining this process:

> *"Consider an organism that must forage in an environment that*

contains edible and inedible objects which are distinguishable on the basis of a number of cues. If two organisms explore this environment side by side, competing for each item encountered, then clearly the one that can make a decision more rapidly as to edibility or inedibility will be able to scoop up the edible objects first and thereby gain a competitive advantage. The organism with a faster decision strategy will high a higher rate of energy intake, and thus will be at an advantage, for example accrue enough energy to produce offspring sooner."

It is uncertain whether the ability to maximize the distance needed to travel between available foods sources yielded the same evolutionary benefits, human performance in the inverted TSP might indicate that this has not been the case.

According to Polavinova (1974, cited in MacGregor and Chu (2011)) TSP problem solving is "firstly based on the general appearance of the itinerary (considering such features as convexity, smoothness, the presence of obtuse angles, the absence of crossing lines, simplicity, aesthetic appearance) and secondly on complexes of points related by their degree of proximity to each other." In other words, the decision maker uses the shape or the environmental structure of the puzzle to guide the decision process. (Simon, 1955) pioneered the research of looking into the interplay between the human mind and the surrounding environment, and with the theory of bounded rationality he suggested that when making a decision, the cognition engage in a search process that is stopped [decision made] when the first alternative that satisfies an aspiration level is met (Gigerenzer, 2000). By aspiration level, he meant the "value of a goal variable that must be reached or surpassed by a satisfactory decision alternative that may go up and down depending on the time spent searching" Simon (1957, cited in Gigerenzer, (2000)). This search can furthermore be divided of into satisfying: search for decision alternatives (the choice set) and fast and frugal heuristics (search for cues in the environment) (Gigerenzer, 2000). If the environmental structures used in the decision process are uninformative to the problem at hand, using them might lead to considerable bias or error.

Tversky and Kahneman (1974) initiated the research of systematic errors and bias in human decision making and observed that respondents' judgments in a multitude of settings were influenced by arbitrary stimulus. In one of their experiments, they observed that respondents estimating the number of African member nations in the UN, was significantly altered after exposure to random numbers drawn from a wheel of fortune. They concluded that:

> *"In many situations, people make estimates by starting from an initial value that is adjusted to yield the final answer. The initial value, or starting-point may be suggested by the formulation of the problem, or it may be the result of a partial computation. In either case, adjustments are typically insufficient. That is, different starting-points yield different estimates, which are biased towards the initial values. We call this phenomenon anchoring."*

In a recent revision Kahneman and Frederick (2002) described anchoring as an attribute substitution where the "target attribute" is substituted by a different "heuristic attribute" rather than being a result of incomplete adjustment from an initial stimuli. They suggest that a decision-maker intending to make a judgmental assessment initiates a search for a feasible value that can mediate the decision. In instances were values are readily available from for example memory, e.g a subjects own age, search may terminated immediately after recall and a decision is made. In situations where the cognition is unable to provide an immediate response, the subject might engage in an extended search that can include attributes unrelated to the judgment (Kahneman and Frederick, 2002). If the target attributes are substituted by attributes unrelated to the decision they may influence the subjects assessments and lead to bias.

2.4 Research question and hypothesis

The ability to adapt and make decisions fast and frugally by using cues from the surrounding environment have enabled decision-makers across species and generations to gain an evolutionary edge over their competitors. Natural selection have promoted the development of specialized cognitive systems that efficiently solve tasks important to survival and reproduction in the natural world. If these specialized decision tools are used outside their intended domains, they may mislead the decision maker and cause biased and erroneous decisions. When researchers are constructing surveys, they implicitly set up artificial environments with structures that may trigger the usage of such tools. If these structures are designed in a way that supplies participating respondents with uninformative cues that unintentionally gets picked up and used to adjust and inform their judgments. The consequence might be that the results and conclusions rendered from that study might be biased by the survey design. The following hypothesis is that the starting point of the lead-time in the LT-TTO might be used as a heuristic attribute by respondents completing the preference elicitation surveys.

To answer the question: "Does the 'Lead-time' component in the LT-TTO function as a heuristic reference point? " We conducted a LT-TTO study that investigated the effect of the lead-time by varying the starting points in the LT-TTO valuation tasks, the goal was to map patterns in the LT-TTO values that could reveal the presence, magnitude or direction of the hypothesized influence.

3 Data Collection and Methods

A representative sample of the Norwegian population aged 18-85 was invited by e-mail to attend the web-survey used in this study. The survey was run by Synnovate, a global market research company as a part of a PHD-project organized by the Center for Health Services Research at Health Region South-East, and the Department of Health Management and Health Economics at the University of Oslo. The sample was drawn from Synnovate's web-panel consisting of 60 000 respondents who participates in an incentive program arranged by Synnovate. Both fixed and lottery-based incentives are provided to motivate the panel member's participation. The incentives are described by the company as moderate and are devised to make incentives appear as nice and motivational, but not as crucial for the respondents (Synnovate, 2011).

3.1 The EuroQol descriptive system

The health states were described to the respondents using the EuroQol (EQ-5D) system (table 1, table 2) (Gudex, 2005). The EQ-5D was developed by the EuroQoL-group to measure Health Related Quality of Life (HRQoL) to be used addition to more detailed measures of health-related quality of life for increased commensurability (Williams, 2005). The EQ-5D is composed of five broad dimensions selected to encapsule a broad array of symptoms and consists of: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The degree of disability (severity) for each dimension was categorized into three levels: No problems, some or moderate problems, or extreme problems.

Table 1: EQ-5D descriptive system

Regression Model	
Mobility	
1.	No problems in walking about
2.	Some problems in walking about
3.	Confined to bed
Self-care	
1.	No problems with self-care
2.	Some problems washing or dressing self
3.	Unable to wash or dress self
Usual activities	
1.	No problems with performing usual activities (e.g. work, study, housework, family or leisure activities)
2.	Some problems with performing usual activities
3.	Unable to perform usual activities
Pain/Discomfort	
1.	No pain or discomfort
2.	Moderate pain or discomfort
3.	Extreme pain or discomfort
Anxiety/Depression	
1.	Not anxious or depressed
2.	Moderately anxious or depressed
3.	Extremely anxious or depressed

The result is a descriptive system that covers $3^5 = 243$ unique health states (245 including instant death and unconsciousness) that is identifiable

Table 2: EQ-5D health states used in the survey

Health State	EQ-5D	Description[2]
1	11211	No probelms walking No problems with self care Some problems with performing usual activities No pain or discomfort Not anxious or depressed
2	11312	No problems walking No problems with self care Unable to perform usual activities No pain or discomfort Moderately anxious or depressed
3	22222	Some problem walking about Some problems washing or dressing self Some problems with performingusual activities Moderate pain or discomfort Moderately anxious or depressed
4	11113	No problems walking No problems with self care No problems with to performing usual activities No pain or discomfort Extremely anxious or depressed
5	32211	Confined to bed Some problems washing or dressing self Unable to performe usual activities No pain or discomfort Not anxious or depressed
6	21232	Some problem walking abou No problems with self care Some problems with performingusual activities Extreme pain or discomfort Moderately anxious or depressed
7	32233	Confined to bed Some problems washing or dressing self Some problems with performingusual activities Extreme pain or discomfort Extremely anxious or depressed
8	33333	Confined to bed Unable to wash or dress self Unable to perform usual activities Extreme pain or discomfort Extremely anxious or depressed

by a five digit descriptor which specify the level severity in the respective dimensions (11111 for perfect health and 33333 for the worst possible state) (EuroQol group). The generic nature of the instrument creates a "convenient way of collecting descriptive data about HRQoL, and about people's own self-rating of their current health state" (Williams, 2005). This also makes the system "extremely valuable in a QALY-type context" (Williams, 2005) where it is used to generate social values (tariffs) for different health outcomes (Gudex, 2005). The social values are elicited using utility search instruments such as the Standard Gamble (SG), the TTO or the Visual Analogous Scale (VAS).

3.2 The survey

The respondents first provided background information including: gender, age, geographic location, educational level, household income, experiences with severe/chronic disease (personal, close relations, nursing), place of birth, smoking status and number of children.

Then they rated their own health, first using the EQ-5D descriptive system, and secondly on a VAS. Then they were presented with eight EQ-5D health states used in the survey (including perfect health and instant death) and ranked them from best to worst. Thereafter they rated the eight health states (including perfect health and instant death) on a VAS-scale that was anchored by the best and worst health state indicated in the former ranking exercise. Then they completed a simplified TTO valuation of the eight health states. At this point, the respondents were informed that they had completed fifty percent of the survey and that they could take a five to ten minute break before continuing with the second part. They were also informed that if they completed the remaining fifty percent of the survey, they would be included in a lottery were one gift card of NOK 10 000 and two gift cards of NOK 5 000 were announced as prices. After the "break" the respondents completed a psychometric profile and answered questions concerning religion, life after death and viewpoints on euthanasia. Then they completed the LT-TTO valuations were the starting point of the LT-TTO valuation task was manipulated.

Three randomization schemes were in play in the survey, each on the level at individual respondents:

1. The order of presentation for the eight target health states was randomized for the simplified TTO task.
2. The order of presentation was randomized for the complete LT-TTO-task, similar to 1, but separate randomization.
3. The initial length of life A was randomized on the level of individual respondents, so that each respondent had the same starting point for all eight EQ-5D health states to be valued

3.3 Structure of the LT-TTO part of the survey

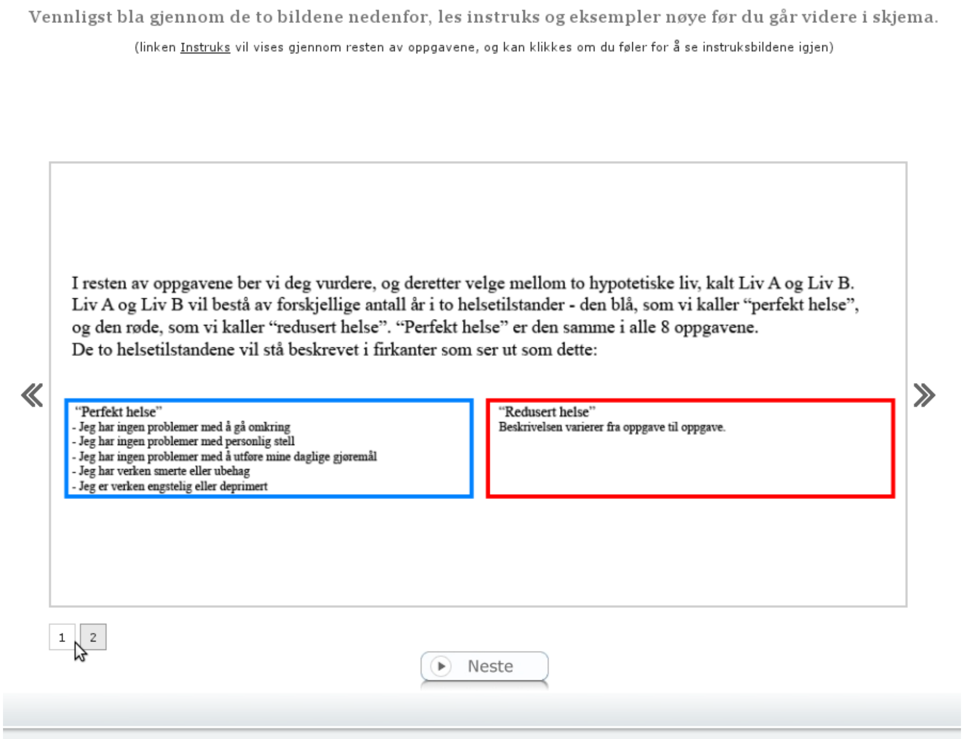
The objective of manipulating the LT-TTO starting point, was to investigate if the point of preferential indifference between two lives would change accordingly. Life B was held constant, while the length of life A was altered until preferential indifference between the two lives was reached. In total, 11 different starting points were used, from 10 to 20 years of perfect health, by increments of 1 year. The LT-TTO part of the survey, was organized in three sections. Section one contained the initial instructions, section two contained the target health state descriptions, while section three contained the LT-TTO valuation task.

3.3.1 Initial instructions

In this section information[3] about how to perform the valuation tasks were provided (fig. 5). This information was also available trough a link on the top of each page throughout the survey. The respondents were informed that they were to value and then choose between two hypothetical lives called life A and life B. They were told that life A would have a specified number of

[3]The survey was conducted in Norwegian, the descriptions provided in the following chapters are translated to English by the author

Figure 5: Initial instructions

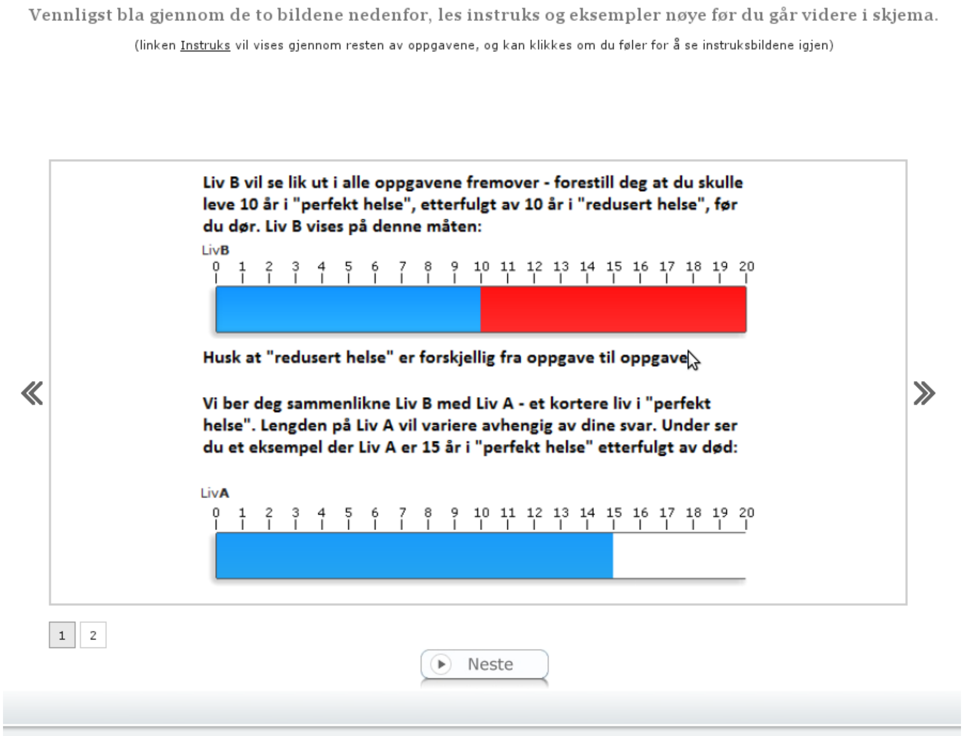


years in perfect health followed by death, while B would have ten years in perfect health followed ten years in the target health state followed by death. The years in perfect health were depicted in blue while the target state were depicted as red. They were informed that "perfect health" would be the same in all eight tasks. A commented example of the LT-TTO sliders were also provided in this section (fig. 6).

3.3.2 Target health state descriptions

Section two and three were both parts of the valuation task. Section two contained the EQ-5D health state description of the state to be valued in section three (fig. 7). The first box (with blue borders) contained a description of the state perfect health (11111), while the second box (with red borders), contained the description of one of the target health states. The respondents were encouraged to "take a good look on the two health states"

Figure 6: Description of how to conduct the valuation tasks

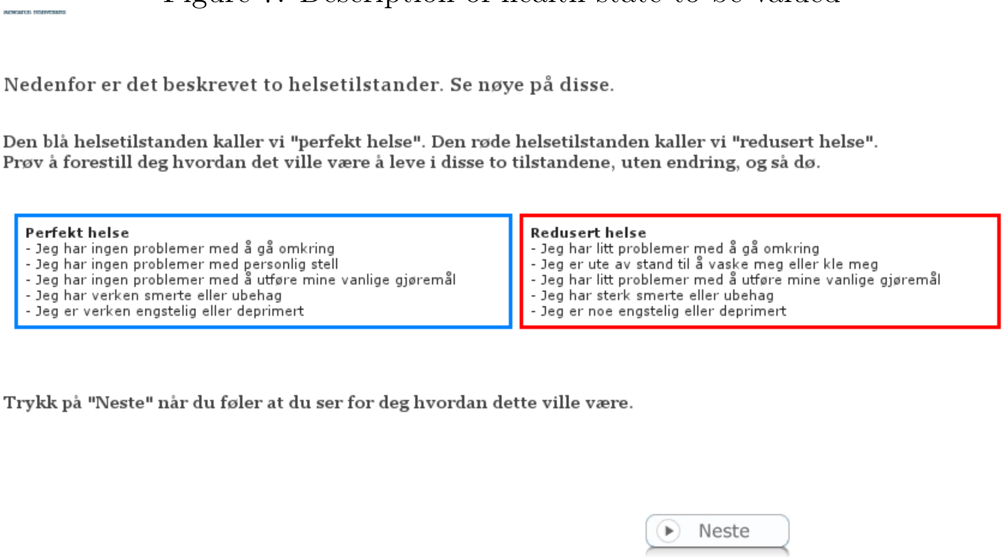


and instructed to "imagine" themselves as living in these two states and then die. When "they felt how that would be" they were instructed to proceed by clicking on the 'next' button".

3.3.3 LT-TTO choice tasks

In the LT-TTO exercise (fig. 8) the respondents were encouraged to imagine how it would be to live in either life A or Life B, and to assess which life they would prefer. The length of life A was defined by the starting point randomization and varied between 10-20 years in perfect health before dying. Life B had a fixed length always consisting of ten years in perfect health, followed by ten years in the target health state followed by death. The respondents were then given three options to specify their preferences: (i) Prefer life A (ii) Prefer life B, and (iii) Both states are equal. The respondents communicated their preferences to the web-survey by clicking on a button

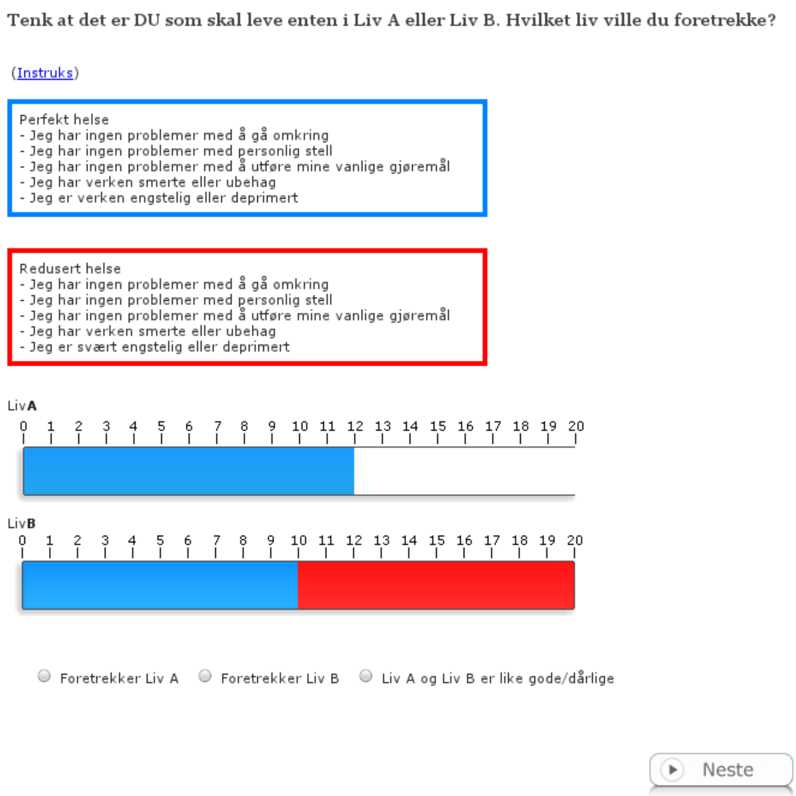
Figure 7: Description of health state to be valued



labeled ”next“. If a respondent preferred life A, and then clicked on the ”next button“, the blue LT-TTO sliding bar (indicating the length of life A) would be reduced by a year. The ”preferring life B“ option would cause the same movements on the sliding bare as in ”preferring life A“ with the exception of the LT-TTO sliding bar increasing by a year.This process of sliding bar movement, would continue until the respondent, (1) opted for the ”both states are equal“ and clicked on the *next* button, (2) activated the reverse iteration counter, or (3) exhausted the LT-TTO sliding bar's range of movement. It is unclear whether the respondents were truly indifferent between the states if they were transfered to the next valuation task by (2) or (3) since they—as opposed to (1) —did not involve an active choice.

To activate the reverse iteration counter, the respondent had to make a reversal from the direction of the initial iteration. If for instance, the respondent first opted for ”preferring life A“, clicked on the ”next button“, and then changed mind and opted for ”preferring life B“ and clicked the ”next button“; the sliding bar's incremental change per click would be halved, and represent 0.5 years. If the respondent decided to make yet another iteration (in any direction), the sliding bar would move the quarter of an iteration (0.25 years), and then automatically transfer the respondent to the valuation task.

Figure 8: LT-TTO choice task



If the LT-TTO sliding bar's range of movement was exhausted, the respondent were automatically transfered to the next valuation task. To exhaust the range of movement, the respondent would have to put the sliding bar in a position that was lower than 0 years in perfect health, or above 20 years of perfect health. The former would indicate that a respondent would prefer spending less than 0 years in perfect health contrasted to spending ten years in full health followed by the target state. The latter would imply that the respondent prefer spending ten years in perfect health, followed by ten years in reduced health in contrast to living twenty years in perfect health.

3.4 Methods

Preliminary descriptive analyses were performed to investigate the properties of the data set and to inform the search for statistical methods satisfying the theoretical constraints set by its characteristics. The assessment was conducted by visual assessment and the Shapiro-Wilk statistical test for normality. An ordinary least squares regression analysis (OLS) was undertaken to investigate the relationship between the dependent and independent variables. The model was specified with the LT-TTO values as the dependent variable. The starting year variable, the EQ-5D health states, the randomization position numbers and the demographic variables constituted the independent variables (table 3).

To detect possible violations of the assumptions required for an unbiased OLS regression, we performed both post regression diagnostic tests and visual assessments of residual plots. The testing included the Shapiro-wilk tests for residual normality, residual-versus-fitted plot graphs and Breusch-Pagan tests for heteroscedasticity, variance inflation factor test to detect multicollinearity between the independent variables, a linktes for assessing model specification, and visual assessments of linearity. Based on the results from the post regression tests, the standard OLS was rejected since several of it's assumptions was violated. To deal with the unmet assumptions, regression designs developed to circumvent the violations of the standard OLS were tested. This included weighted robust regression to reduce influence of leveraged outliers, regression with Huber/White/Sandwich variance-covariance estimators to correct issues with heteroscedasticity. Clustered regression were applied to investigate suspicions of dependency related to the variables, Respondent ID, TTO randomization and Health state severity. A trade-off between the consequences of not correcting for the violation of each singular assumption were conducted, and a robust regression with Huber and Tukey bi weights was chosen on the basis of this having the most pronounced effect on the coefficient of the starting year variable.

Post regression diagnostics of the results from the first standard OLS regression we conducted indicated that the assumptions of homogeneity of

Table 3: Regression variables

Variable	**Definition**	**Description**
Starting_years	Starting years	Variable ranging from 0-10 containing the length of the" lead-time".
IStateNum	EQ-5D health state	Dummy-variable for each of the eight different EQ-5D health states.
IRandom	Randomization order	Dummy-variable for the randomization position of the observation.
Iage	Age group	Dummy variable for the three age groups as specified in table 4.
Ichildren	Dummy variable for children	Containing the two categories for Yes/No.
Igeo	Geographic region	Dummy variable for the four national regions as specified in table 4.
Icivil	Marital status	Dummy variable for the subjects' marital status as specified in table 4.
Iwork	Working status	Dummy variable of the subjects' current working situation as specified in table 4.
Iincome	Income	Dummy variable for the differnt household income levels as specified in table 4
Iutd	Education level	The subjects' education levels by years as specified in table 4.
_cons	Constant	The regression constant.

variance and multicollinearity were breached in the data. The set also had a high density of leveraged outliers, and there were as mentioned above, empirical reasons to believe that the assumption of independence was not entirely fulfilled. Concerning model specification, we chose not to omit variables suspected to cause multicollinearity. The decision was based on a risk/reward perspective were the effect of including or omitting variables was weighted against each other. The dilemma consisted of the risk of getting inflated confidence intervals by including variables that proxy each other, contrasted to the risk of bias in the model coefficients caused by omitting important information. Since the confidence intervals for the impact of starting years in the LT-TTO exercise remained robust with different model specifications, we chose to include all the varibles as specified in table 3.

Observations with preference profiles that were considered as implausible or illogical was excluded using four exclusion criteria. (1) Respondents who had rated seven or eight health states as worse than or equal to death. (2) Respondents who rated all states as equal. (3) Respondents who rated the best EQ-5D health state (11211) as worse than the presumed worst health state (33333). (4) Respondents subjected to a minor task error.

3.4.1 Software

Data were analyzed using STATA/SE version 10.1 for Unix based systems. LaTeX were used for typographic processing. STATA output and results were transformed into LaTeX code by using Gnumeric spreadsheet software and the STATA add-ons Tabout, Sjlatex and Outtex.

4 Results

4.1 Study population

The study population (table 4) was predominantly middle aged 59.5 % with a mean age of 44.4 years were 54.6 % of the respondents were male and 46.8% reported to be married or cohabitant, 66.7 % were childless, 56.0 % had at

Table 4: Survey population

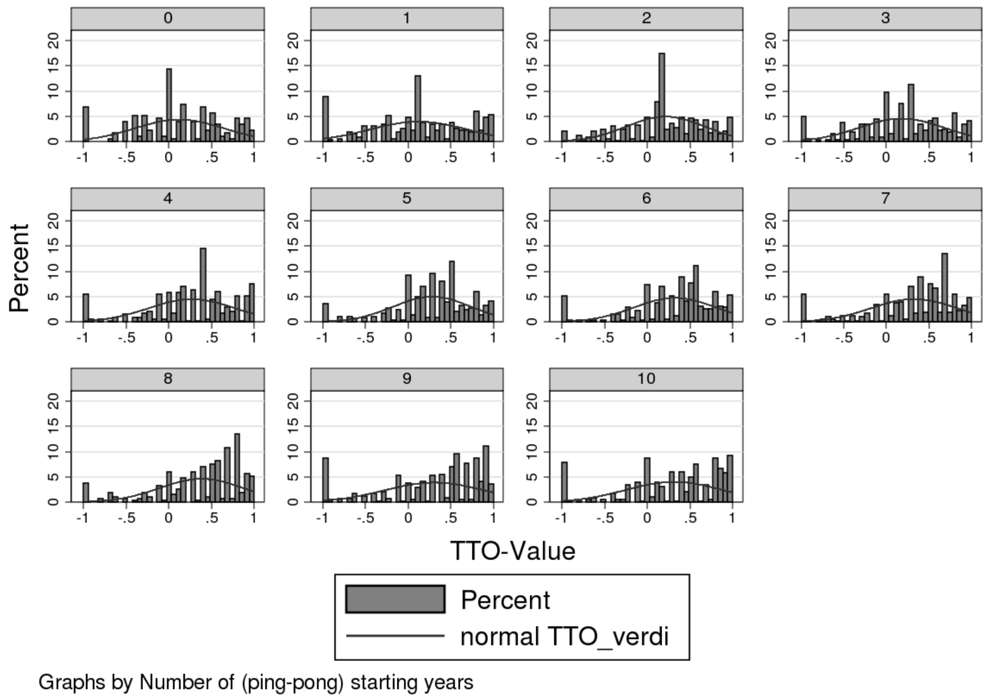
		Norwegian Population	Sample Before Exclusion	Sample After Excluded	Percent Excluded
Sex					
	Female	49.9%	45.9%	45%	20.5%
	Male	50.1%	54.1%	55%	18.9%
Education					
	Less than 8 years	.	1.0%	1%	41.7%
	9-10 years	29.8%	7.1%	7%	18.1%
	11-13 years	42.9%	25.6%	23%	29.1%
	> 12 Years	27.3%	56.8%	59%	16.1%
	Under education	.	9.5%	10%	14.4%
Income					
	NOK 0-100.000		1.9%	2.0%	18.2%
	NOK 100.-199.000		2.8%	2.9%	17.2%
	NOK 200.-299.000		5.5%	4.6%	32.7%
	NOK 300.-399.000		12.7%	12.0%	24.0%
	NOK 400.-499.000		12.2%	11.9%	21.4%
	NOK 500.-599.000		11.6%	10.9%	24.4%
	NOK 600.-799.000		20.8%	22.1%	14.7%
	NOK 800.-999.000		12.9%	13.5%	15.6%
	NOK 1 mill. +		11.1%	11.5%	16.7%
	Don't want to answer		4.1%	4.0%	21.7%
	Don't Know		4.4%	4.6%	17,0%
Age					
	18 - 30 years	19%	22.2%	24.2%	12.6%
	31 - 60 years	41%	59.4%	59.5%	19.6%
	61 - 85 years	19%	18.3%	16.3%	28.6%
Children					
	No		66.7%	65.2%	21.4%
	Yes		33.3%	34.8%	16.2%
Geographic Region					
	East-Norway		55.2%	54.9%	20.2%
	West-Norway		28.0%	28.2%	19.2%
	Mid-Norway		12.4%	12.7%	17.7%
	North-Norway		4.4%	4.3%	22.0%
Working Status					
	Full time		59.8%	59.8%	19.7%
	Part time		7.8%	7.9%	18.8%
	Self Employed		3.7%	4.1%	10.1%
	Pension		7.4%	6.9%	25.0%
	Unemplyed		1.2%	1.0%	32.1%
	Trygd		6.5%	6.0%	25.7%
	Student		9.7%	10.6%	12.3%
	Home worker		0.2%	0,2%	0,0%
	Other		3.7%	3.4%	25.6%
Marital Status					
	Married/Partner		46.9%	46.8%	19.9%
	Co-living		22.1%	22.1%	19.5%
	Unmarried/Single		20.3%	20.2%	20.0%
	Seperaated / divorced		10.8%	11.0%	18.2%

Table 5: Median and mean TTO values for the different starting year groups

	Median	**Mean**	**Min - Max**
EQ-5D Health States			
1 (n=461)	0.82	0.75	$[-1, 1]$
2 (n=457)	0.50	0.46	$[-1, 1]$
3 (n=450)	0.50	0.45	$[-1, 1]$
4 (n=450)	0.38	0.29	$[-1, 1]$
5 (n=459)	0.40	0.34	$[-1, 1]$
6 (n=457)	0.15	0.07	$[-1, 1]$
7 (n=456)	0.10	0.02	$[-1, 1]$
8 (n=454)	-0.12	-0.27	$[-1, 1]$
Total (n=3,644)	0.34	0.27	$[-1, 1]$
Rotasjon TTO 1			
1 (n=434)	0.34	0.24	$[-1, 1]$
2 (n=455)	0.30	0.25	$[-1, 1]$
3 (n=457)	0.30	0.25	$[-1, 1]$
4 (n=459)	0.35	0.26	$[-1, 1]$
5 (n=461)	0.33	0.25	$[-1, 1]$
6 (n=460)	0.30	0.26	$[-1, 1]$
7 (n=459)	0.38	0.32	$[-1, 1]$
8 (n=459)	0.40	0.29	$[-1, 1]$
Total (n=3,644)	0.34	0.27	$[-1, 1]$
Number of (ping-pong) starting years			
0 (n=175)	0.12	0.12	$[-1, 1]$
1 (n=315)	0.10	0.12	$[-1, 1]$
2 (n=327)	0.20	0.21	$[-1, 1]$
3 (n=318)	0.20	0.18	$[-1, 1]$
4 (n=330)	0.38	0.27	$[-1, 1]$
5 (n=436)	0.31	0.28	$[-1, 1]$
6 (n=314)	0.40	0.32	$[-1, 1]$
7 (n=401)	0.42	0.33	$[-1, 1]$
8 (n=267)	0.50	0.38	$[-1, 1]$
9 (n=415)	0.50	0.31	$[-1, 1]$
10 (n=346)	0.40	0.32	$[-1, 1]$
Total (n=3,644)	0.34	0.27	$[-1, 1]$
Total (n=3,644)	0.34	0.27	$[-1, 1]$

Source: Synnovate

Figure 9: Histogram TTO-values by starting years



least 12 years of education, 59.8 % was employed full time, and 54.9 % lived in eastern Norway.

The exclusion criteria excluded 19.7 % of the total number of respondents (table 6). The exclusion criterion "best state worse or equal to death" contributed with 17.3 % of the exclusions, "all states rated as equal" excluded (6.0 %), "all states WTD" excluded (9.7 %) and the "technical error in registration" criterion excluded (17.5 %) of the excluded.

The respondents were mostly evenly distributed across the three randomization schemes, providing a balanced demographic profile to most of the testing variables. The exclusion criteria, however, discriminated certain demographic profiles and lead to a higher exclusion rate for some sub-groups. The groups with the highest rate of exclusions were the respondents with less than eight years of education (41.7 % exclusion), the respondents receiving NOK 200.-299.000 in yearly income (32.7%), and the unemployed (32.1%) exclusions. There was also variation in proportions of excluded across the dif-

Table 6: Excluded % in the different starting year groups

	Starting Years											
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Tot.**
best_state_under_worst												
0 (n=3,752)	61.3	71.4	82.1	87.0	78.2	87.6	81.9	88.1	83.4	91.7	89.9	82.7
1 (n=784)	38.7	28.6	17.9	13.0	21.8	12.4	18.1	11.9	16.6	8.3	10.1	17.3
Total (n=4,536)												
Pearson chi2(10)=	203.8 Pr=		0.000									
similar_values												
0 (n=4,264)	80.8	87.5	96.1	97.8	89.1	96.9	96.0	96.5	95.3	98.3	96.1	94.0
1 (n=272)	19.2	12.5	3.9	2.2	10.9	3.1	4.0	3.5	4.8	1.7	4.0	6.0
Total (n=4,536)												
Pearson chi2(10)=	89.80		Pr=	0.000								
all_WETD												
0 (n=4,096)	69.5	75.0	92.7	93.5	89.1	90.8	90.0	96.5	92.9	100.0	98.0	90.3
1 (n=440)	30.5	25.0	7.8	6.5	10.9	9.2	10.1	3.5	7.1	0.0	2.0	9.7
Total (n=4,536)												
Pearson chi2(10)=	371.5 Pr=		0.000									
task_error												
0 (n=4,415)	96.6	98.0	96.8	98.6	96.4	98.0	97.0	98.5	96.4	97.0	96.8	97.3
1 (n=121)	3.4	2.0	3.3	1.4	3.6	2.0	3.0	1.5	3.6	3.0	3.2	2.7
Total (n=4,536)												
Pearson chi2(10)=	11.00 Pr=		0.357									

Source: Synnovate

ferent starting-year groups. (Pearson χ^2 (10 d.g) = 170.3 p-value = 0.0000). The starting-year group 0 which initially received only 6.4 % of the respondents (optimal proportion would be $\frac{1}{11} = 9.1\%$), had an exclusion rate of 40 % which is twice the number of excluded compared to the groups (2-10). Here, all had an exclusion rate of less than 21.1 %.

4.1.1 Exclusions

In group 0: 38 % of the respondents rated seven or eight health states as worse than or equal to death, 19.2 % rated all states as equal (indifferent to the health state severity), and 30.5 % rated the best EQ-5D health state (11211) as worse than the presumed worst health state (33333).

4.1.2 Hypothesis test

The KruskalâWallis test detected significant differences between the populations chi-squared = 222.306 with 10 d.f. probability = 0.0001.

4.2 Regression model

The regression analysis of the LT-TTO values (table 7) suggest that an incremental increase of the starting points inflate the mean TTO-values by a factor of 0.039 (Prob > F = 0.0000), and indicate that the respondents populating starting year group 10 arrives at a 39% mark-up on their TTO mean values as compared to their peers in the lowest lead time starting group. The dummy variables for the randomization order were also statistically significant, adding evidence to the claim that the order of the LT-TTO tasks may influence the valuations. The dummy varibles representing the severity of the valued states were also statistically significant, this was expected since they initially were selected by the property of being differentiable. Several of the demographic variables were also significant.[4]

The Shapiro-Wilk tests for normality and visual analysis of the residual plot confirmed that the residuals were normally distributed. A leverage-versus-squared-residual plot indicated a pronounced presence of leveraged outliers. Heteroscedasticity were confirmed by the Breusch-Pagan LM statistic: 1191.513 Chi-sq(41) P-value = 0.0000 and by visual confirmation of a residual-versus-fitted plot. The variance-inflation-factor test (VIF) detected multicollinearity between the independent variables, this was specially prominent in the education and income status variables were some of the dummies had a VIF of 20.69 (_Iutd_4), 11.29 (_Income_2) and 10.26 (_Income_1).[5] The linktest _hat = 1.00, p= 0.000, _hatsq = -0.03 p = 0.068 indicated that relevant variables were not ommited. Visual assessments of linearity confirmed that the starting-point had a linear relationship with TTO-values in the tested intervals.

[4]However, we recommend not putting to much effort into the interpretation of the significance-levels of the demographics since they were sensitive to multicollinearity

[5]A cut-off point of 10 are normally considered as a an indication of multicollinearity.

Table 7: Regression model.

Robust Regression with Huber- and Tukey Bisquare -weights

TTO_value	**Coef.**	**Std. Err.**	**t**	**P>t**	**[95% Conf.**	**Interval]**
Starting_years	.0390728	.0024785	15.76	0.000	.0342138	.0439319
_IStateNum_2	-.265745	.0300474	-8.84	0.000	-.3246526	-.2068374
_IStateNum_3	-.2757266	.0300607	-9.17	0.000	-.3346604	-.2167927
_IStateNum_4	-.4199225	.0300666	-13.97	0.000	-.4788679	-.360977
_IStateNum_5	-.3650164	.0300463	-12.15	0.000	-.4239219	-.306111
_IStateNum_6	-.6224805	.0300463	-20.72	0.000	-.6813861	-.563575
_IStateNum_7	-.67417	.0300682	-22.42	0.000	-.7331185	-.6152214
_IStateNum_8	-.9504569	.03006	-31.62	0.000	-1.009389	-.8915245
_IRandom_2	.0646361	.0300615	2.15	0.032	.0057007	.1235715
_IRandom_3	.082728	.0300686	2.75	0.006	.0237787	.1416772
_IRandom_4	.0954335	.0300595	3.17	0.002	.0365021	.1543648
_IRandom_5	.0879994	.0300575	2.93	0.003	.0290719	.1469268
_IRandom_6	.0826663	.0300508	2.75	0.006	.0237519	.1415806
_IRandom_7	.1263658	.030046	4.21	0.000	.0674608	.1852708
_IRandom_8	.1010251	.0300384	3.36	0.001	.042135	.1599151
_Iage_1	-.0915451	.0269432	-3.40	0.001	-.144367	-.0387232
_Iage_2	-.1709274	.0364153	-4.69	0.000	-.2423194	-.0995354
_Ichildren_1	.0852356	.018758	4.54	0.000	.0484607	.1220105
_Igeo_1	.0495999	.0177785	2.79	0.005	.0147454	.0844544
_Igeo_2	.004004	.0243828	0.16	0.870	-.0437983	.0518063
_Igeo_3	-.0176266	.0376517	-0.47	0.640	-.0914425	.0561892
_Icivil_2	-.0710363	.0217117	-3.27	0.001	-.1136019	-.0284707
_Icivil_3	-.0694384	.0256604	-2.71	0.007	-.1197455	-.0191314

continued...

_Icivil_4	-.0114964	.0268576	-0.43	0.669	-.0641505	.0411578
_Iwork_2	-.0817423	.0300802	-2.72	0.007	-.1407144	-.0227703
_Iwork_3	-.0255445	.0411785	-0.62	0.535	-.1062746	.0551857
_Iwork_4	-.0128271	.0385169	-0.33	0.739	-.0883392	.062685
_Iwork_5	-.2506272	.0711411	-3.52	0.000	-.3900987	-.1111556
_Iwork_6	.0231551	.0334815	0.69	0.489	-.042485	.0887953
_Iwork_7	-.0051153	.0339115	-0.15	0.880	-.0715986	.0613679
_Iwork_8	-.0874398	.1864556	-0.47	0.639	-.4529844	.2781048
_Iwork_9	-.0906085	.0423328	-2.14	0.032	-.1736015	-.0076154
_Iincome_1	.0126914	.0400423	0.32	0.751	-.0658111	.0911939
_Iincome_2	.0409971	.0416043	0.99	0.324	-.0405678	.1225621
_Iincome_3	-.087305	.0536785	-1.63	0.104	-.1925412	.0179313
_Iincome_4	.0931874	.0520758	1.79	0.074	-.0089069	.1952816
_Iutd_2	-.0290522	.08172	-0.36	0.722	-.1892637	.1311592
_Iutd_3	-.1057208	.0790988	-1.34	0.181	-.2607934	.0493518
_Iutd_4	-.0443145	.0785598	-0.56	0.573	-.1983304	.1097013
_Iutd_5	-.0720944	.0834751	-0.86	0.388	-.2357466	.0915578
_cons	.5017628	.0946614	5.30	0.000	.3161799	.6873457

5 Discussion

The purpose of this thesis was to investigate whether the starting point in the LT-TTO could affect the elicited results. We hypothesized that the starting year would influence the elicitation process by serving as a heuristic anchor. Evidence supporting the claim were detected by statistical analysis and regression modeling. The differences in the TTO values between the starting-year groups had a positive linear relationship between the starting point and the TTO-values.

Post regression diagnostics indicated that some of the assumptions required for an unbiased OLS regression analysis were violated in the data. This included the requirements of independent observations and constant variance of the error terms. There where also issues relating to influential observations and multicollinearity. Since (a) the respondents rated eight subsequent health states (Dolan, 1997), and (b) the severity of previously rated health states might affect the valuations of the following health states; the error terms might be dependent on the individual respondent valuing the health states and the health state valuation order might cause correlations between the error terms of the different health states. Independence is a requirement for the significance tests used in the OLS regression, if the assumption is not met the significance levels of the independent variables might be biased. Heteroscedastistic variance of the error terms might also influence the significance level and bias their inferences. The presence of multicollinearity might affect both confidence intervals and regression coefficients, to correct for this we fitted a regression model were the "high" VIF variables (working status and level of education) and observed that it caused no change in the coefficients of the relevant variables. It nonetheless affected some of the confidence intervals of the demographic variables and increased the number of significant variables. The Huber and Tukey bi weight regression model was the only design that had a marked effect on the coefficients of the starting year groups.

Since the significance levels of the relevant variables remained in the p=0.0000 range for all model specifications and regression models correcting for the observed violations, the bi weighted regression model seemed like the best approach to the data.

5.1 Discriminating exclusion criteria

An unexpected peculiarity was the between group discrimination of the respondents in starting year group 0 by the exclusion criteria. The exclusions caused by the "best state under the worst" and "all states rated as worse than or equal to death" might be attributed to the group being positioned closest to the negative values. Albeit this does not explain why the group also had a higher propensity to not differentiate (similar values) between the eight health states. Another explanation might be that there is some property, not present in the higher starting years, that increase the probability of the respondent misunderstanding how to conduct the valuations. If the high exclusion rates in this group is caused by the latter explanation, the implication might be that there are some unexplored dynamics relating to the starting point of the lead time that affects the respondents interpretation and understanding of the valuation task.

5.2 Heuristic attributes

There are two kinds of attributes that stands forth as potential candidates for causing the variance observed in the data. The first is the heuristic anchor following the tradition of Tversky and Kahneman, where an individuals perception of something is affected by an arbitrary factor. The second type can be coined as "environmental attributes" and lie in proximity to Herbert Simon's theory of bounded rationality. The environmental attributes can be seen as non-heuristic properties that influence individuals' judgments' trough enforcements of technical paradigms or environmental structures. From this point on, we will therefore draw a distinction between "heuristic attributes" and "environmental attributes".

5.2.1 Heuristic attributes in the graphical representation

To answer a LT-TTO survey a respondent need to engage in several cognitive tasks. First, the semantic meanings of the EQ-5D descriptions need to be interpreted. Second, the target health state has to be conceptualized, either by drawing on experience or by the usage of imagination. Third, the target health state then needs to be compared and assessed relative to perfect health. Fourth, the relative assessment needs to be translated into a time scale. Fifth, the respondent's valuation of the hypothetical health states measured in years need to be communicated to the system by the aid of a graphical representation and a digital interface.[6]

The cognitive process needed to solve each of the five LT-TTO judgmental tasks probably possess enough material for years of scientific inquiry. The main interest of this study revolves around stages four and five. The reason is that in these two stages the respondent translates and communicates the perceived relative values of the target health states, and transform them from cognitive concepts to ratios expressed by a time-scale. Unless the respondent has a clear perception of his or her preferences for the hypothetical health states expressed in a health-time continuum, it is reasonable to believe that the respondent initiates a search for cues in the environmental structures of the LT-TTO. Because the level of abstraction needed to process and articulate the valuation, the target attribute might be inaccessible for the respondent. Since the initial ratio of the graphical bars expressed by the starting-year might provide a readily available candidate for substitution, the cognitive system might accept the starting year as a reference point (heuristic attribute) and substitute it for target attribute. Incomplete adjustments away from the heuristic attribute might influence the respondent's valuations, pull them in the direction of the heuristic attribute, and influence the respondent revealed preferences for the target health state as observed in the data.

5.2.2 Technical attributes in the digital interface

Changing the position of the lead-time will naturally alter the distance between the initial position of the sliding-bar and the point of preferential indifference. Consider the example of a respondent in starting-year group 10, if he or she were indifferent between perfect health and spending time in the target health state. The respondent would have to move the sliding bar from the starting point (at ten years) to the end point (at twenty years). In comparison, a respondent in group 20 could communicate indifference by leaving the sliding-bar at it's initial position. The survey-design used in this study opens for two ways of measuring the distance between the starting

[6]It is therefore not given that the results are transferable to an analogous setting.

points and the target values. First, "as the crow flies" by the number of iterations expressed by the sliding bar. Second, "as the man walks" indicating the number of mouse-clicks needed to reach the point. Since the respondent have to click on two buttons to achieve a single iteration (i+1) on the sliding bar, "the real walking distance" is twice as long as the distance "the crow flies" $(2i + 1)$. Group 10 would for instance reach a QALY-value of either 1 or -1 with ten iterations or twenty mouse clicks in either direction. Group 5 on the other hand, would need fifteen clicks to reach a value of 1, and twenty-five clicks to reach -1, Group 10 would need zero clicks to reach a QALY-value of 1, and forty mouse clicks to value a state as -1. This creates an unbalanced distribution of distances between starting- and the end-point of the LT-TTO valuation task and might serve as a technical attribute that influences the respondents' valuations.

If respondents are sensitive to the distances between the starting-point and the latent point of preferential indifference, it would logically follow that the starting-points in the LT-TTO would influence the respondent's revealed point of preferential indifference. A theoretical justification of the potential sensitivity can be found in our treatment of the TSP in chapter two. We hypothesized that the human performance in solving TSP-problems could be related to the cognitive architecture developed trough evolutionary processes that favored energy conserving behavior. Moreover, the need to make fast and frugal decisions promoted the development of heuristic tools that search the environment for cues to mediate a judgment. Additionally, the search was thought to be guided by a "stopping-rule" contingent on adaptive aspiration levels searching for satisfactory outcomes.

If we imagine a person in group 20, that is infinitely sensitive to negligible reductions in health, meaning that he or she would assign the value -1 to all health states less than perfect. The person would need to click the mouse button 41 times in order reach the point of preferential indifference for every health state. The total amount of time this respondent would spend on just clicking (assuming that the two clicks would take about 1 second) would amount to $8 * 41 = 328$ seconds or approximately $5\frac{1}{2}$ minutes. In the other extreme situation, where an individual in group 20 is unwilling to trade quality with length of life, assigning all health states the value 1. He or she would only have to click on the *next* button to finish the task, meaning that the total number of clicks would amount 8, and by our assumption, have a total clicking time of 8 seconds, which is 41 times lower than the phobic respondent. The point is not to make an accurate postulation about the time it takes to push the two buttons in the interface, rather, to give an example of how a digital interface might provide a structure that in first eyesight are equal for all groups or personalities, but after some consideration might be a significant source of variance.

If human cognition uses the same tools in artificially constructed environments in natural environments, there is reason to believe that the same pattern of behavior emerges. Increasing the effort of reaching an accuracy level may for instance convince an individual to accept a less preferred (accurate) result as satisfying because the effort of reaching the optimal increase. To make a modern example, imagine two local food stores in close vicinity of your home. They are equally accessible, but you perceive one of them as slightly better than the other. Which one would you visit when you are out of merchandise? Now, imagine that your preferred store moved and the distance you need to travel to it increased by an inconvenient factor, would it affect? Unless the preferred store was very special, most people would probably settle for the simpler store for convenience. Applying the same logic to the artificial structure of the digital interface, the suggestion is that some of the variation in and between the starting-year groups is a result of aspiration levels adapting to unbalanced distribution of the distances.

5.3 Limitations

Some of the assumptions required for OLS regression were clearly unmet in the data, statistical methods designed to circumvent the violations found little or no differences in the study-relevant variables when corrected for. Moreover, a non-parametric statistical test added corroboration to the findings by detecting a significant between-group variation. Despite of this, scientific rigor dictates that the the model should be considered as biased due to the failure of assumption adherence. The violations reduces the validity of the findings and a careful approach to the interpretation of the regression output is therefore stressed.

5.3.1 Weaknesses in the survey design

The greatest weakness of this study was the failed implementation of the "ping-pong" debiasing method (Lenert et al., 1998) . Miscommunication between the researchers and the programming staff resulted in an algorithm that, instead of providing the ping-pong method, created a "reversal iteration counter" that counted the number of iterations (i_r) if a respondent switched direction of the initial TTO-sliding bar movement, the code instructed the sliding-bar to behave in the following way

$$i_r = 1 \rightarrow i = \frac{1}{2}i \tag{9}$$

and in the next iteration after the reversed iteration

$$i_r = 2 \vee i_r = 1 \rightarrow i = \frac{1}{4}i \tag{10}$$

and after the second iteration after the reversed iteration

$$i_r = 3 \vee i = 1 \rightarrow NEXT \tag{11}$$

Said in other words, for $i < 3$ iterations after a reversed iteration, the iteration-steps is reduced by $\frac{1}{2}$ for each iteration. When $i = 3$ after a reversed iteration, the program ends the current valuation task and automatically transfer the respondent to the next task. Presumably, the respondents were unaware of the algorithms, and the indicated point of preferential indifference might be a function an automatic transfer rather than a conscious decision. A regression model run solely on the respondents subject to the "automatic transfer" (46% of the observations) indicated that they had a slightly lower coefficient (0.033 [0.025 0.042] p = 0.0000) attributed to the starting years than the group not subjected to this error.

Another weakness also related to the programming concerns another "automatic transfer". If the LT-TTO sliding-bar was exhausted, meaning that the slider sits in a position that would yield a QALY value of either 1 or -1, and the respondent opted for an extra iteration in the exhausting direction. The respondent would automatically be taken to the next valuation tasks.

$$h_\beta > 1 \vee h_\beta > -1 \rightarrow NEXT \tag{12}$$

This "magnet effect" pulled in total x respondents on the negative side and y respondents on the positive range of the TTO-values. The starting year group 0 had X and the starting year group 10 had Y. The other groups all had Z. If the respondent experienced the disruption as an unintentionally event, the transfer might bias the results.

6 Conclusion

The between-group variation in this study of LT-TTO valuation suggests that the respondents' observed willingness to sacrifice life duration for increases in health quality, depends on the researchers decision of the lead-time's initial starting points. The observed variation raises doubts whether the LT-TTO leads us closer to an unbiased preference elicitation tool, or if it only trades one type of heuristic effect off for another. Since it is difficult to perceive that the dependency is a result of an *a priori* property of peoples' perceptions of health and time, anchoring should be considered when designing and interpreting LT-TTO studies.

References

Dan Ariely, G. Loewenstein, and D. Prelec. Arbitrarily coherent preferences. *The psychology of economic decisions*, page 131, 2004.

E. Chronicle, J. MacGregor, and T. Ormerod. Optimizing and âpessimizingâ: Human performance with instructional variants of the traveling salesperson problem. *The Journal of Problem Solving*, 1(1):7, 2006.

L. Cosmides and J. Tooby. Better than rational: Evolutionary psychology and the invisible hand. *The American Economic Review*, 84(2):327–332, 1994.

N.J. Devlin, A. Tsuchiya, K. Buckingham, and C. Tilling. A uniform time trade off method for states better and worse than dead: feasibility study of the "lead time" approach. *Health Economics*, 20(3):348–361, 2011. ISSN 1099-1050.

P. Dolan. Modeling valuations for EuroQol health states. *Medical care*, 35 (11):1095–1108, 1997. ISSN 0025-7079.

P. Dolan, C. Gudex, P. Kind, and A. Williams. The time trade-off method: Results from a general population study. *Health Economics*, 5(2):141–154, 1996.

M.F. Drummond, M.J. Sculpher, and G.W. Torrance. *Methods for the economic evaluation of health care programmes.* Oxford University Press, USA, 2005. ISBN 0198529457.

M. Dry, M.D. Lee, D. Vickers, and P. Hughes. Human performance on visually presented traveling salesperson problems with varying numbers of nodes. *Journal of Problem Solving*, 1(1):20–32, 2006.

G. Gigerenzer. *Adaptive thinking: Rationality in the real world. 2000.* Oxford University Press, Oxford, UK.

S.M. Graham, A. Joshi, and Z. Pizlo. The traveling salesman problem: A hierarchical model. *Memory & cognition*, 28(7):1191–1204, 2000.

C. Gudex. The descriptive system of the EuroQol Instrument. *EQ-5D concepts and methods: A developmental history*, pages 19–27, 2005.

D. Kahneman and S. Frederick. Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The psychology of intuitive judgment*, pages 49–81, 2002.

L.A. Lenert, D.J. Cher, M.K. Goldstein, M.R. Bergen, and A. Garber. The effect of search procedures on utility elicitations. *Medical Decision Making*, 18(1):76, 1998. ISSN 0272-989X.

J.N. MacGregor and Y. Chu. Human performance on the traveling salesman and related problems: A review. *The Journal of Problem Solving*, 3(2):2, 2011.

C.H. Papadimitriou. *Computational complexity*. John Wiley and Sons Ltd., 2003.

A. Robinson and A. Spencer. Exploring challenges to TTO utilities: valuing states worse than dead. *Health economics*, 15(4):393–402, 2006. ISSN 1099-1050.

H.A. Simon. A behavioral model of rational choice. *The quarterly journal of economics*, 69(1):99, 1955.

P.F.M. Stalmeier, J.J.V. Busschbach, L.M. Lamers, and P.F.M. Krabbe. The gap effect: discontinuities of preferences around dead. *Health economics*, 14(7):679–685, 2005. ISSN 1099-1050.

C. Tilling, N. Devlin, A. Tsuchiya, and K. Buckingham. Protocols for time tradeoff valuations of health states worse than dead: a literature review. *Medical Decision Making*, 30(5):610, 2010. ISSN 0272-989X.

P.M. Todd. Fast and frugal heuristics for environmentally bounded minds. *Bounded rationality: The adaptive toolbox*, pages 51–70, 2001.

G.W. Torrance, W.H. Thomas, and D.L. Sackett. A utility maximization model for evaluation of health care programs. *Health services research*, 7 (2):118, 1972.

G.W. Torrance, M.H. Boyle, and P. Horwood. Application of multi-attribute utility theory to measure social preferences for health states. *Operations Research*, 30(6):1043–1069, 1982. ISSN 0030-364X.

A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124, 1974. ISSN 0036-8075.

I. Van Rooij, A. Schactman, H. Kadlec, and U. Stege. Perceptual or Analytical Processing? Evidence from Childrenâs and Adultâs Performance on the Euclidean Traveling Salesperson Problem. *The Journal of Problem Solving*, 1(1):6, 2006. ISSN 1932-6246.

H.R. Varian. *Intermediate microeconomics: a modern approach*. WW Norton, 2006. ISBN 0393927024.

D. Vickers, M. Butavicius, M. Lee, and A. Medvedev. Human performance on visually presented traveling salesman problems. *Psychological Research*, 65(1):34–45, 2001.

J. Von Neumann and 0. Morgenstern. Theory of games and economic behavior. *Princeton University Press*, 1947:1953, 1944.

A. Williams. The EuroQol Instrument. *EQ-5D concepts and methods: A developmental history*, pages 1–17, 2005.

R. Zeckhauser and D. Shepard. Where now for saving lives? *Law and contemporary problems*, 40(4):5–45, 1976.