The London School of Economics and Political Science

The Measurement of Quality-Adjusted Life Years: Investigations into Trade-Offs Between Longevity and Quality of Life

Sarah M. Watters

A thesis submitted to the Department of Social Policy at the London School of Economics for the degree of Doctor of Philosophy, London, August 2016.

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 66,771 words.

Statement of use of third party for editorial help

I can confirm that my thesis was copy edited for conventions of language, spelling and grammar by James Watters.

Acknowledgment

To my family, thank you, more than you could ever know. And to Adam Oliver, for challenging me and for inspiring me to think more deeply.

Abstract

In health care, decision makers are faced with increasing innovation and demand for services accompanied by escalating costs. As a result, governments and institutions have sought to promote health care value (i.e. better outcomes per moneys spent). A summary measure of health-related quality-of-life (HRQoL) to help decide how to allocate available resources is thus highly desirable. In no other area of public policy has a measure similar to the widely-used quality-adjusted life-year (QALY) been developed. The QALY is therefore unique in both its ambitions and in the political, philosophical and measurement challenges it faces.

This thesis set out to examine health state valuation using the time-trade off (TTO), a tool used to measure HRQoL, in the context of a behavioural economic framework. Observed violations of procedural and descriptive invariance, cornerstones of decision theory (on which the TTO is based), have been witnessed in health state valuation and elsewhere. Behavioural economics offers a framework by which such inconsistencies can potentially be better understood. Although behavioural economics has gained traction in other areas of decision research, its application to health state valuation has been limited.

Drawing on the decision-making literature and health-specific considerations, the empirical studies in this thesis: provide insight into why previous studies of the TTO have yielded inconsistent findings, showcase violations of internal consistency due to behavioural economic phenomena, and identify issues relevant to the choice of TTO ‘version’ (i.e. *how* values should be elicited). Implications of the research in terms of stated preference methods and their role in policy are discussed.

A strict focus on the TTO was intended, as it is the tool most widely implemented in health state preference elicitation, both in research contexts and clinical studies that seek to demonstrate cost-effectiveness. However, importantly, the empirical findings and discussion in this thesis are relevant not only to researchers of health state valuation but to policy makers in health and other areas of social policy which seek input for their decisions through stated preference exercises.

Table of Contents

List of Tables

List of Figures

List of Boxes

List of Abbreviations

CBA	Cost-Benefit Analysis
CEA	Cost-Effectiveness Analysis
CPTO	Constant Proportional Trade-Off
CUA	Cost-Utility Analysis
DCE	Discrete Choice Experiment
EQ-5D	EuroQol 5-Dimensional Questionnaire
FH	Full Health
HRQoL	Health-Related Quality of Life
HS	Health State
LE	Life Expectancy
LSE	London School of Economics and Political Science
MCDA	Multiple Criteria Decision-making Analysis
NICE	National Institute for Health and Care Excellence
QALY	Quality-Adjusted Life Year
SLE	Subjective Life Expectancy
SWB	Subjective Well-Being
TTO	Time Trade-Off
VAS	Visual Analogue Scale

Chapter 1: Introduction

1.1 Introduction

1.11 The Allocation Problem

It has long been acknowledged that health is a fundamental component of both societal and individual welfare (Grossman, 1972). On an individual level, Bentham (1780) wrote of the *relief of pain* as a basic or simple pleasure. On a broader scale, health [care] allows individuals to be fully participating citizens who are able to contribute to the social, political, and economic life of the society in which they live (Daniels, 2001). Thus, it is in the best interest of society to have a health care system that functions effectively.

In health care, as in virtually all public sectors, resources are limited and must be rationed. Available resources are facing rising pressure due to medical innovation, changes in both patient and carer demands, and demographic trends involving higher rates of age-related diseases and chronic conditions (Cracknell, 2010; Dormont et al., 2006; Dormont and Huber, 2006; Lafortune et al., 2013). Health care policy makers have a responsibility to allocate the resources available to them in a way that contributes to societal welfare through improved health. Ultimately, deciding which interventions to provide (and to whom) is an ever-present issue at both clinical and policy levels and the methods guiding how resources might best be distributed are central to addressing the allocation problem.

Broadly, at the policy level, health care resource allocation encompasses two main (frequently competing) objectives: to allocate with efficiency (i.e. extract maximum benefits from resources available) and to allocate with equity (i.e. distribute the benefits as equitably as possible) – the equity-efficiency trade-off (Brock, 2003; Wagstaff, 1991). Improvements in health and well-being are often cited as a moral concern in that the absence of health may compromise the individual's ability to flourish (Culyer, 1992; 2006). Incorporating equity considerations (e.g. prioritizing by severity of the condition), however, may result in an efficiency loss, in that more benefits could have been produced by a different allocation. Thus, the tension in balancing efficiency and equity considerations constitutes an important ethical debate (Brock and Wikler, 2006). At issue in this thesis are the methods which underpin the efficient allocation of resources in health care, specifically, how they may fall short in adequately informing such decisions.

The manner in which health care resources are allocated is notably different than in other areas of social policy. This is primarily due to the fact that the value of benefits in health care cannot be directly assessed through market behaviour (Arrow, 1963). Although there are other areas of social policy in which traditional market structures are not present (e.g. education and environmental policy), additional problems arise in determining a desirable allocation of health care resources (Robinson and Hammitt, 2011). For instance, an individual does not typically enter the health care market by choice. Rather than seeking health care, in the sense that it is a discretionary commodity, individuals demand health care as a means of obtaining good health (Grossman, 1972). The individual also does not directly face the full cost of their health care at the time of consumption (since in the UK, for example, health care is financed through taxation). Moreover, informational asymmetries (e.g. where the individual has little information as to the benefit of a given treatment compared to the provider) and uncertainty with regards to the nature and timing of health care consumption divorce the individual from the market[i] (Arrow, 1963; Haubrich and Wolff, 2006; Hurley, 2000).

Due to such market failures, economic evaluation has become a bastion in health care resource allocation, providing a perspective from which efficient allocations can be estimated in the face of otherwise indeterminate demand (Johnson and Bingham, 2001). As a means of guiding policy decisions, economic evaluation helps to inform which health care interventions should be available to the public by examining the opportunity costs and outcomes of allocation alternatives (Arrow, 1963; Drummond, 2005). The utilitarian modus operandi of economic evaluation in health care is to maximize health within a given set of health care resources.

Health systems in Australia, Canada, and Norway, inter alia, rely on tools of economic evaluation (CADTH, 2006; PBAC, 2013; NOMA, 2012). In England and Wales, the National Institute for Health and Care Excellence (NICE) was established in 1999 to offer guidance as to whether interventions should be made available to the public. NICE seeks to inform what constitutes *value for money* – i.e. an efficient use of the resources available to the NHS and personal social services (Rawlins and Culyer, 2004).

1.12 Defining the Numeraire: Valuing Health

At its most elementary level, economic evaluation requires the comparison of inputs and outputs in order to determine which interventions offer the greatest health gain per amount spent (i.e. the greatest output from the fewest inputs). Three types of economic evaluation are most relevant to health care: cost-benefit analysis, cost-effectiveness analysis, and cost-utility analysis. While each expresses costs (inputs) in monetary terms, the quantification of health benefit (outputs) differs.

In cost-benefit analysis (CBA) - the traditional welfare economic approach to valuing benefits - both costs and benefits are expressed in monetary terms (Drummond, 2005; Mishan, 1971). CBA is frequently applied in other areas of social policy such as transport and environmental economics (Arrow et al., 1996; Freeman, 1979). While several strategies have been put forward to derive a monetary value for the benefits of an intervention one commonly used technique is that of contingent valuation.[ii] Contingent valuation creates a hypothetical marketplace, requiring an individual to state their willingness to pay (or to accept) for the benefits (or disadvantages) associated with a particular outcome (Arrow et al., 1993).

The decision rule in CBA is that social welfare improves if net benefits (e.g. health gains) exceed net costs. Given that both the inputs and outputs are monetized, CBA is able to determine the net social benefit of an intervention without a comparison intervention (Drummond et al. 2005). CBA has the ability to address questions relating to both technical and allocative efficiency. Technical efficiency refers to achieving minimum input for a given output. On the other hand, allocative efficiency refers to the arrangement (provision of resources) where maximum benefits are derived from a given set of resources. Notably, CBA can be used to compare inputs and outputs not only within a health context but also other areas of interest within the same budget constraint. Notwithstanding its advantages, as Garber and Phelps (1997, p. 28) comment "the [monetary] valuation requirement for CBA is both its greatest disadvantage and its greatest strength." Concerns have been put forward relating to the ethical nature of monetizing health or an individual's life. Additionally, scope insensitivity, whereby an individual's willingness to pay values lack sensitivity to differences in the number of units under consideration (i.e. the size of the outcome), has been shown to inhibit accurate measurement (Gyrd-Hansen et al., 2012; Olsen et al., 2004; Shiell and Gold, 2003). As this thesis will show, however, as a method of economic evaluation, CBA is not alone in its measurement problems.

Cost-effectiveness analysis (CEA) is an alternative to CBA that offers a comparative analysis of interventions using a common output measure. CEA enables different interventions to be compared based on a common outcome, such as blood pressure, glucose levels, or life-years gained, therefore addressing questions relating to technical but not allocative efficiency. CEA is thus useful when comparisons must be made within a condition-specific group (i.e. where there is a common indicator of interest).

In CEA (as well as in cost-utility analysis, presented below), a monetary value is not explicitly assigned to health outcomes (benefits). Instead, the unit of analysis is the incremental cost-effectiveness ratio (ICER), the ratio of resource input to health output. An ICER compares the intervention of interest with a comparator intervention, typically the most cost-effective alternative. To calculate an ICER, the difference between the costs of the two interventions (where C2 and C1 are the costs of the intervention of interest and its comparator, respectively) is divided by the difference in effects (where E2 and E1 are the effects of the intervention and its comparator, respectively). The lower the ICER, the greater is the output (i.e. health gain) from a given input (i.e. resources, moneys spent) and thus the more cost-effective the intervention.

In CEA (as well as in cost-utility analysis), the decision rule for economic evaluation is seen in equation (1) where v^i represents the consumption value (a value judgment, further discussed below in relevance to cost-utility analysis) and costs and effects are expressed by ΔC and ΔE respectively. If the ICER falls below v^i then the intervention is considered cost-effective:

$$(1) \qquad \Delta C / \Delta E < v^i$$

An important shortcoming of standard CEA is the inability to make comparisons between health programs or interventions where the outcomes are different, hence precluding CEA’s usefulness in comparing across disease areas, for example. Cost-utility analysis (CUA) emerged as means by which effects (ΔE) can be operationalized so as to incorporate both morbidity (i.e. non-fatal health outcomes) and mortality in a composite outcome, the quality-adjusted life-year (QALY). The QALY merges length of life (measured in life years) and health-related quality of life (hereon HRQoL, measured on a number of dimensions including social, physical, and emotional functioning) into an index of value-weighted time (i.e. life-years weighted by HRQoL).

As a “common currency” (p. 39), the QALY enables comparisons across widely varying disease areas by indexing all health states on a single continuum, assigning each state a *health state value* (McKie et al., 1996). Health state values are typically scored on a scale of 0 to 1, where 1 implies full (‘perfect’) health and 0 corresponds to death. The lower the value the less desirable the health state is considered to be. Negative values are possible and imply that the given state is considered to be worse than death.

To illustrate the basic principles of QALY calculations, consider the following simplified example illustrated in Figure 1-1. Suppose an individual has a health condition (e.g. hypertension) assigned a value of 0.40 and a life expectancy of 5 years. With treatment, their life expectancy is increased to 10 years and their HRQoL to a health state value of 0.7. The QALY gain from the intervention, calculated as the difference between no treatment (0.4*5 years = 2 QALYs; area C) and the treatment scenario (0.7*10 years = 10 QALYs; area A+ B), is 5 QALYs, assuming no time discounting (further discussed in Section 1.53). In Figure 1-1, area A and area B represent the gain in HRQoL and the improvement in longevity, respectively.

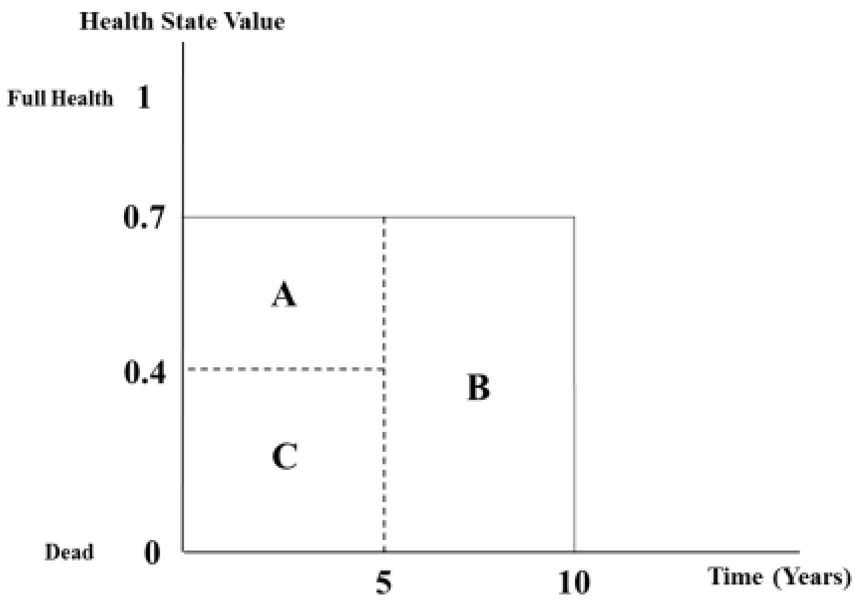


Figure 1-1: An example of QALYs gained from an intervention (area A + B) compared to no intervention (area C)

That preferences are cardinal is central to health state value measurement since this allows judgments of magnitude in addition to order, permitting meaningful comparisons to be made (Coons and Kaplan, 2005). For cardinality to be satisfied, it is necessary that a gain in health, for example, from 0.4 to 0.6 is equal to a gain from 0.6 to 0.8 and that a health state assigned a value of, say, 0.6 is considered to be three times better than a health state with a value of 0.2 (Bossert, 1991). In order for health state values to be legitimately aggregated to inform social policy, an important assumption is that they are interpersonally comparable. This means that values can be compared between individuals and that every person's life has the same value (i.e. the relative value of QALY gains or losses to different individuals is equal).

1.13 Allocating Resources Using CUA

CUA is used by NICE in England and Wales to provide guidance on the cost-effectiveness of new health technologies. CUA in itself is a descriptive exercise, meaning unlike CBA, where the cost-effectiveness of a single intervention can be assessed in isolation (i.e. by calculating whether monetary expenditures exceed monetary benefits), CUA requires an external benchmark (a value judgment, v^i from equation (1)) from which the intervention of interest can be deemed worthwhile (Menzel et al., 1999). It is generally accepted that NICE uses a benchmark in the form of a cost-effectiveness threshold ranging between £20,000-30,000 per QALY.[iii] Interventions with ICERs that fall below the threshold are normally recommended whereas interventions with ICERs that exceed the threshold are typically not recommended (with exceptions, for example, in the case of end-of-life circumstances) (Claxton et al., 2013). Recently, as evidence of the significance of such a threshold, Dakin et al. (2014) found that a given intervention's ICER could predict 82% of NICE's decisions.

Cookson and Culyer (2010) comment that deviations from efficiency objectives (i.e. basing decisions solely on cost-effectiveness) are both desirable and necessary, an idea shared by many others (e.g. Hausman, 2010; Henshall and Schuller, 2013). To this end, an equity-weighted consumption value may be established. For example, exceptions to NICE's cost-effectiveness threshold have been made in instances where it has been decided that other factors should be incorporated (NICE, 2014). However, while there is considerable support for incorporating distributional concerns into resource allocation, researchers and policy-makers face challenges in reaching consensus in terms of the most important considerations and how they should be operationalized into allocation decisions. Thus, determining what ought to be the elements of an ethically sensitive QALY remains a work in progress (Shah et al., 2013).

Regardless of the distributional desideratum, in order for CUA to be an admissible policy tool, a number of decisions must be made in terms of the various ICER inputs and their analysis. The decision maker must decide, for example, which perspective on costs and outcomes is to be taken (i.e. which costs are relevant and should be included), how costs and outcomes (QALYs) should be discounted, and how to handle uncertainty

(Brouwer and Koopmanschap, 2000; Briggs and O'Brien, 2001; Claxton et al., 2011; NICE, 2013). Fundamental to CUA analysis, and the issue which is addressed in this thesis, is the derivation of health state values.

The remainder of this introductory chapter considers the elicitation of health state values. The next section briefly reviews *how* health states are described in valuation tasks and *who* is to value them, followed by a discussion of three commonly used elicitation methods, with emphasis on the time trade-off, the method focused on in this thesis. In Section 4 important problems with the TTO's internal consistency are raised, and a framework in which these inconsistencies can be interpreted is laid out in Section 5. Section 6 introduces possible ways forward in addressing TTO inconsistencies, the specific approach taken in this thesis, and the rationale for this approach. An outline for the remaining chapters in the thesis then follows.

1.2 QALYs

The QALY's raison d'être – to enable comparisons of varying types of health outcomes using a single index – is unrivaled by any other outcome measure. The notion of having a single index by which otherwise disparate health outcomes can be compared is highly desirable. In seeking to capture such a desirable measure some degree of simplification and abstraction is necessary. The question then arises as to whether the simplifications are so substantial that they inhibit proper measurement of the health states and thus undermine the validity of the cardinal values from which the QALY is built.

In the absence of readily observable preferences in a real market (i.e. revealed preferences), in order to attach values to different health states, health economists have opted for stated preference methods that depend on hypothetical choices as the best alternative solution. As we will see, although the QALY numeraire is used widely, the methods used to assign health state values are subject to much criticism, particularly when they are considered in light of the findings from the wider judgment and decision-making literature.

The next section offers a brief introduction to some of the key considerations in the elicitation of health state values. Three questions relating to health state value measurement, proposed by Dolan (2000), are addressed: "what is to be valued; who is to value it; and how is it to be valued." A fourth question, how are differences in values to be dealt with, has elicited differing responses which include averaging value differences and relying on deliberative methods. The first two questions are addressed briefly and the fourth question lies beyond the scope of the material covered in this thesis. The focus of the thesis is on the third question, specifically, how well current methods capture the construct of interest, HRQoL. An outline of health state valuation is followed by a discussion of the central question of health state validity (i.e. how well health state valuation methods represent HRQoL) and of the behavioural economic concepts which form the analytical framework for the research in this thesis.

1.21 The Description of Health

Health states can be described in several ways for health state value elicitation. Common methods include vignettes, generic classification systems and disease and condition specific frameworks. Vignettes take the form of text narratives or a more structured bullet point format. Usually tailored to a specific condition, vignettes can incorporate a range of information pertaining to the symptoms and treatment (including side-effects) (Brazier and Rowen, 2011).

A number of generic, multi-attribute health state classification systems (MAUs) offer comparability across disparate health states by classifying health states according to particular dimensions. The EuroQol Descriptive System (EQ5D-5L), for example, is preferred by NICE and comprises five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) with five levels on each dimension (no problems, slight problems, moderate problems, severe problems and unable to/extreme problems).[iv] Other commonly used MAU instruments include the Short Form 6D Health Status Questionnaire (SF-6D) (Brazier et al., 2002), the Health Utilities Index Mark 3 (HUI-3) (Furlong et al., 2001), and the 15-D measure (Sintonen, 2001). There is variation in the dimensions of health included in each instrument and the different levels within each of the dimensions so it is therefore unsurprising that the results of the different instruments have been shown to be inconsistent (Kopec and Willison, 2003; McDonough et al., 2007; Seymour et al., 2010).

MAU descriptive systems may obscure important differences in levels of severity for a given condition or fail to capture particular disease-specific impacts, and thus it is preferable in some instances to describe the health state using disease- or condition-specific terms (Brazier et al., 2007; Revicki and Kaplan, 1993). For example, a condition-specific or domain-specific description is useful if there is a particular symptom or dimension of functioning that is of interest to the researcher (Bowling, 2001).

1.22 Whose Values?

A particular source of contention involves *who* are the most appropriate respondents for health state value elicitations. Experts, such as physicians or other medical personnel, patients, or members of the general public are all possible options. An ongoing debate surrounds (empirically observed) differences in health states values elicited from patient samples and the general population.[v] Ultimately, the question of *whose values should count* remains unresolved, but the use of public preferences is currently the dominant perspective in CUA and recommended in the NICE reference case (which outlines a core set of methods developed to promote consistency in economic evaluations) (NICE, 2013). Support for this perspective is found in arguments which propose that these respondents are acting under a veil of ignorance (and thus do not take their own future health or self-interest into consideration), as well as on the basis of general public accountability - i.e. the general public should have a role in setting priorities for the health care they are receiving (De Wit et al., 2000; Menzel, 1999; Whitty et al., 2008). To this extent, eliciting general population preferences can be seen as “a natural extension of the principles of democratic governance” (Gregory et al., 1997; Payne et al., 1999, p. 243).

1.23 The Valuation of Health

Health economists have a number of tools available to elicit health state values. This section provides an introduction to these tools, focusing largely on the time trade-off, the method examined in this thesis.

Among the most frequently used methods to elicit health state values are rating scale methods (category rating and the visual analogue scale - VAS), the standard gamble and the time trade-off.[vi] As choice-based methods, the standard gamble and time trade-off have been traditionally favoured among economists. They represent compensatory models which require respondents to forgo one attribute for an improvement in another and thus can be grounded in economic (particularly, utility) theory (Chapman et al., 1999). On the other hand, when using rating scale methods, the value is simply the point where the health state is placed on the given scale and can therefore be termed a "choiceless utility” (Gold et al., 1996; Loomes and Sudgen, 1983, p. 428). The choice of method used to elicit health state values is important since empirical evidence has found they produce different values for the same health state[vii] (e.g. Bleichrodt and Johannesson, 1997; Stiggelbout et al., 1994).

The VAS and category rating tasks ask the respondent to assign the health state to a point on a given scale, often from 0 (death or least desirable) to 10, or 100 (full health or most desirable). While arguably easier to understand than choice-based methods, these techniques are susceptible to biases such as end aversion (where ratings tend to converge towards the middle of the scale) (Bleichrodt and Johannesson, 1997; Streiner and Norman, 1989; Torrance et al., 2001).

The standard gamble embodies the notion of risk which some believe is an advantage since health care interventions invariably involve a degree of risk (see Mehrez and Gafni, 1991). In the standard gamble the respondent is presented with two options: a certain outcome or a gamble. The gamble, which would lead to a better quality of life, also poses the risk of death (1-p) while the same life expectancy is provided for both living health states. The respondent is asked to state the probability of success of the gamble (i.e. a probability equivalence) for which they would be indifferent between it and the certain outcome. The probability (p) (i.e. the amount of risk) the respondent is willing to incur reflects their perceived utility of the heath state.[viii]

Among other concerns, respondents may have trouble interpreting the risk element of the standard gamble (e.g. see Buckingham et al., 2004; Ryan et al., 2009) and the time trade-off (TTO) has emerged as an alternative. The trade-off between HRQoL and longevity presented in the TTO is generally consistent with the concept of value as defined in economics, whereby making trade-offs between important attributes or goods[ix] underlies high-quality, rational decision-making (Freeman, 1993; Frisch and Clemen, 1994 as in Payne et al., 1999). The ease with which TTO values can be computed into QALYs is an advantage of the TTO, highlighted by Torrance (2006), who commented that the TTO is essentially a QALY equivalence statement in that it collapses the relationship between time and HRQoL into a single measure (whereas in the standard gamble, for example, the health state value may be confounded by risk attitude) (Richardson, 1994). Further, the TTO has been found to outperform the standard gamble on measures of logical and internal consistency and reliability (Dolan et al., 1996). At present, the TTO is arguably the most prevalent technique for valuing health states in health economics (Boye et al., 2014; Dolan and Roberts, 2002; Wisløff et al., 2014). It is recommended by NICE and the Scottish Medicines Consortium (SMC) in the UK and a number of countries have used the TTO to generate EQ5D tariff sets using general population values (e.g. the Netherlands, Japan, France, Argentina, China, and Denmark) (Augustovski et al., 2009; Chevalier et al., 2011; Devlin et al., 2011; Lamers et al., 2006; Liu et al., 2014; Samuelsen et al., 2012).

In its standard interpretation, the TTO pairs a trading metaphor with a search procedure (also termed response mode in the literature) (Bennett et al., 2002). The TTO asks the respondent to indicate the number of years (*x*) spent in full health they require to be indifferent to *t* years in a lesser health state (i.e. they would trade years to avoid worse health – e.g. *headaches* in Figure 1-2). The resulting TTO value (*x*/*t*) is the value assigned to the health state.

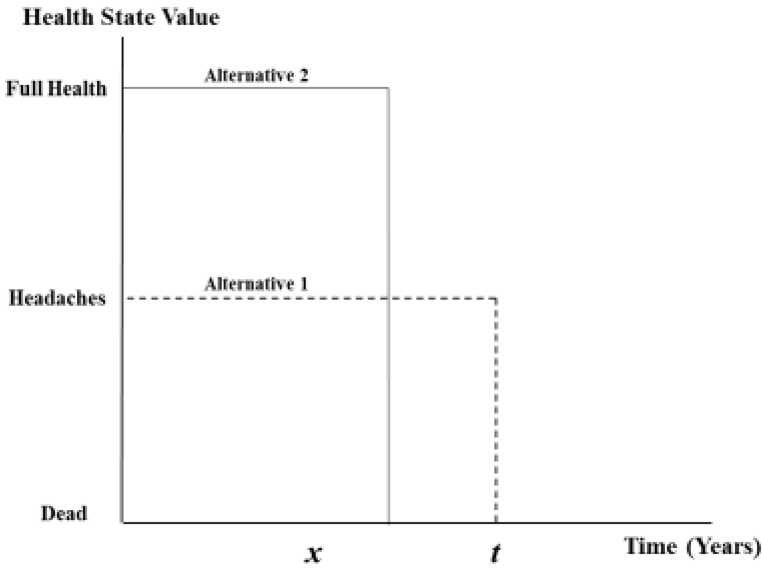


Figure 1-2: The time trade-off (TTO)

A TTO task may ask respondents the following question, “Imagine that you have two treatments available: one treatment will give you *t* (for example, 10 years) with *headaches* (e.g. headaches for an hour every day) and a second treatment will give you *x* (which is smaller than *t*) years in full health (FH). What is the minimum number of years, *x,* in full health which would make you indifferent between (FH, *x*) and (*headaches, t*)?” (*x/t*=value for the health state *headaches*).

Although the TTO has a relatively simple mathematical interpretation, the task can be structured in a number of different ways, as can be seen in Table 1-1. Most notably, trade-offs between HRQoL and longevity may be made from different time horizons. An important, and restrictive, assumption of the TTO model is that of constant proportionality (constant proportional trade-off, CPTO). CPTO implies that the individual assigns the same value to each unit of time regardless of the time horizon, such that the value of each individual year in a 40-year TTO, for example, will be the same as if it was a 10-year TTO.

In this way, if the individual were to indicate indifference at 20 years in full health in the 40-year TTO for health state *x* (health state value=20/40=0.5), they would (proportionally) indicate indifference at 5 years in full health in the 10-year TTO for health state *x* (5/10=0.5). CPTO allows for the value elicited for a given health state to be independent of the time horizon used in the TTO task. On a practical level, it means that values elicited at any given time horizon can be inputted into CUA calculations in which the duration of the health state differs from that used to elicited the health state value.

Table 1-1: Overview of variation in elements of TTO design from select studies

Study	**Life years used**	**Search Procedure**	**MET health state**	**Utility correction**	**TTO long …TTO short**	**Sample**
Sackett and Torrance (1978)	3 months; 8 years; LE	Choice	Mixed	No	<	general population, patients
Miyamoto and Eraker (1988)	1, 2, 15, 16, 20, 24 years	Matching	No	No	<	patients
Stiggelbout et al (1995)	3, 10, 15 years; 3, 5, 10 years; 5 years, 20 years, LE	Matching	No	No	<	patients
Dolan and Stalmeier (2003)	10 and 20 years	Up titration	Yes	No	<	students
Martin et al (2000)	5, 10, 15 years	Matching	No	Yes	<	CVD patients over 60
Stalmeier et al (2001)	10 and 20 years	Choice	Yes	No	<	Students, patients
Essink-Bot et al (2007)	10 years and actuarial LE	Ping-pong	Mixed	No	<	general population
Kirsch and McGuire (2000)	2 and 10 years	Choice/TTO props	Mixed	No	≤	general population
Hall et al	10% of LE; 50% of LE; LE	Ping-pong	Mixed	No	equal	women (ages 40-70); 50% patients
Stalmeier et al (1996)	5, 10, 25, 50 years	Choice	Yes	Yes	equal	students
Bleichrodt and Johannesson (1997)	10 and 30 years	Matching	No	No	equal	students
van der Pol and Roux (2005)	20 and 50 years	Choice	No	Yes	equal	students (n.b. between-subjects design)

Pliskin et al (1980)	5 years; 15 years	Matching	No	Yes	equal (about half were linear in their preferences)	pilot sample (n=10)
Attema and Brouwer (2013)	3, 10, 12, 31, 46 years	Choice/ matching	No	Yes	= (matching) > (choice)	students
Stalmeier et al (1997)	5, 10, 25, 50 years	Choice	Yes	No	≥	students
Unic et al (1998)	5 and 10 years, higher (15-60, depending on remaining LE)	Choice	No	No	>	healthy women
Bleichrodt et al (2003)	13, 19, 24, 31, 38 years	Ping-pong	No	No	≥	students
Rencz et al. (2015)	20 and 80 years	Matching	Mixed	No	≥	young adults with and without migraines
Attema and Brouwer (2008)	14 and 27 years	Matching	No	Yes	>	students
Attema and Brouwer (2010)	10, 20, 40 years	Matching	No	Yes	inverse u-shape	students
Attema and Brouwer (2012)	3, 10, 12, 31, 46 years	Choice	No	Yes	inverse u-shape	students
Lenert et al. (1998)	30 years	Ping-pong	No	No	NA	general population

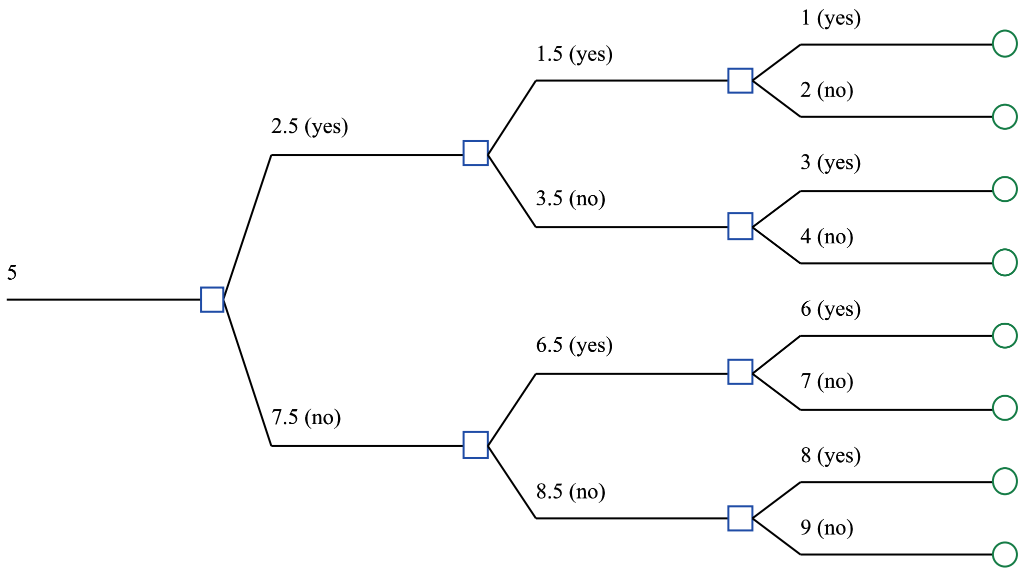
"TTO long…TTO short": direction of difference in TTO values comparing long and short time horizons (life years used)
LE: life expectancy

Another way in which TTO tasks may differ is in the method by which the point of indifference is reached – the search procedure. In order to obtain the TTO value, an indifference point between the number of years spent in full health and the number of years spent in the target health state must be established. While there are a number of strategies by which the indifference point can be reached (obtaining preference orderings), they can be broadly categorized as either choices or judgments (hereon matching). Choice-based methods include bisection, ping-pong and upward and downward titration, as described in Box 1-1.

Box 1-1: Search procedures in the TTO

A *matching* procedure requires the respondent to directly value one alternative in order to achieve indifference (i.e. perceived equivalence) between the options. The task requires the respondent to provide their indifference value between two options without any prompting from the researcher (i.e. they are to 'fill-in-the-blank'). For example, respondents will be asked to state the number of years in full health that they consider to be equal to 10 years with back pain (i.e. without being presented the options of 10, 9, 8, ... or 10, 1, 8…).

Choice-based methods require the respondent to select among alternatives presented to them. Such methods include *upward* and *downward titration* procedures as well as *bisection* techniques. In an upward titration technique, the respondent is asked to make choices with the number of years in good health ascending (e.g. 1 year in full health or 10 in health state *x*, 2 or 10, 3 or 10…). Conversely, in a downward titration, the respondent is presented with outcomes where the number of years in good health is descending (e.g. 9 or 10, 8 or 10, 7 or 10…). In a bisection procedure, the respondent is offered a choice between *x* years in the health state (e.g. back pain) or a percentage – initially determined by the researcher – (e.g. 50% of *x* years, hence *bisection*) in full health. Based on their response to this initial question, the respondent is presented with a series of further choices where the number of years in full health is further bisected until an indifference point is attained. Shown below is an example of the bisection iteration sequence for the years of full health in a TTO offering the (initial) choice of 10 years in the health state or 5 years in full health.



More generally, the TTO question may be posed in other, strategically equivalent, ways, for example by asking the respondent to forgo quality of life for greater longevity (a ‘reverse’ TTO). Recently, a number of investigations examined lead- and lag-time TTOs where a period of full health is added to either the beginning or the end of the period in which trade-offs are made (Devlin et al., 2013).[x] Put simply, a lead-time TTO, for example, might ask the respondent the indicate how many years they would be willing to forgo from a 10 year period with migraines beginning after an initial 5 more years in full health.

In general, to assume that a health state value is a valid representation of an individual’s strength of preference for a given health state, standard decision theoretic principles must hold (implying the individual is rational in their decisions). Invariance, a central tenet of rational decision-making, states that different representations of the same choice problem should elicit the same preference. Therefore, regardless of the specific methods (e.g. time horizon, search procedure) used within any one of the TTO iterations (i.e. standard, reverse, lead- or lag-time) the TTO value should be consistent. More specifically, TTO valuations should be *invariant* regardless of how outcomes are described (descriptive invariance) or the way in which the choice is elicited (procedural invariance). Preferences for health states are also assumed to be complete and transitive. Completeness implies that the individual has a known subjective value (underlying preference) for an outcome and thus all outcomes are comparable: they prefer A to B, prefer B to A, or are indifferent between A and B. If the individual possesses *transitive* preferences, then if they prefer A to B and B to C, they also prefer A to C. Pending such assumptions (invariance, completeness, and transitivity), it can be inferred that the individual possesses a consistent preference ordering for a particular set of outcomes.

1.3 Violations of the Assumptions

An important problem faced by the TTO model (and the QALY model more generally – i.e. regardless of the elicitation method used) is that health state values appear to significantly depend on the decision context. For example, there is evidence that TTO values are dependent on how the health states are described, which search procedure was used, the time horizon, and the other health states being valued in the same study (e.g. Attema and Brouwer, 2012; Bleichrodt et al., 2003; Lenert et al., 1998). Thus, contrary to the assumption that preferences for health states are well-formed, stable and “effectively ‘data’ waiting to be collected” (Shiell et al., 2000, p. 47), values are dependent on the particular circumstances in which they are derived.

TTO inconsistencies are problematic for several reasons. Given that health states represent a psychological concept, no external referent exists against which values can be compared (Richardson, 1994). Also, it is difficult to know which values are the best reflection of underlying preferences and thus should be used to calculate QALYs. A practical implication of inconsistent TTO values for the same health state is that if (value) differences are large enough, different cost-effectiveness or policy decisions may be made depending on which value is used. Borrowing on an example in Oliver and Wolff (2014), Box 1-2 provides an illustration of such an instance using data from a pilot study for this thesis. Further, Box 1-3 (below) outlines TTO design features that may contribute to relatively high or low TTO values for the same health state.

Box 1-2: An example of TTO value inconsistency in cost-effectiveness determinations

To illustrate the effect of inconsistencies in TTO values on cost-effectiveness, consider the following simplified example. Two groups of respondents were asked to state the number of years in full health which would make them indifferent with 10 years with back pain. One group provided their response in an open-ended (matching) format, that is, they simply stated their indifference point. The other group’s indifference point was ascertained by a series of iterative questions which asked the respondent if they would prefer a specified number of years in full health over 10 years with back pain, where the specified number of years in full health was bisected in each step. The mean TTO value in the matching group was 0.84 and 0.73 in the bisection group.

Suppose that for each individual an intervention for back pain generates an additional 10 years of life at a total cost of £225,000. If the matching TTO value of 0.84 is inputted into cost-effectiveness calculations then the intervention will yield 8.4 QALYs (0.84*10). Alternatively, if the TTO value elicited through bisection is inputted then the intervention would yield 7.3 QALYs (0.73*10). Based on a cost-effectiveness threshold of £20,000-30,000/QALY, if the matching value is used, the intervention would be deemed good value for money (£225,000/8.4 QALYs=£26,786/QALY) but would not be considered cost-effective if the value had been elicited using a bisection technique (£225,000/7.3 QALYs = £30,822/QALY) assuming zero discounting.

Box 1-3: Key study design features that would promote "high" TTO values and "low" TTO values for the same health state

Extremely Low TTO Values	Extremely High TTO Values
- Downward titration - Short time horizon - Matching search procedure - Health state the respondent is familiar with and/or considers to be relatively severe - Previous question about a much milder health state - Previous question with a much longer time horizon	- Upward titration - Long time horizon - Health state the respondent is familiar with and/or considers to be relatively manageable - Previous question about a more severe health state - Previous question with a much longer time horizon

TTO values have been shown to be subject to framing effects whereby logically inconsequential differences in how outcomes are described affect the values which are elicited (Tversky and Kahneman, 1986). Rabin (1998, p. 37) explains that framing effects will be observed if different presentations of a problem direct the respondent's attention in varying degrees towards different aspects of the problem. Recent investigations have shown that adding a period of full health before or after the period of time over which years are traded in lead- and lag-time TTOs appears to create a framing effect. Specifically, when longer lead- or lag-times were offered, the more time the respondent was willing to trade-off (Augustovki et al., 2013; Janssen et al., 2013; Oppe et al., 2014).

Framing effects result in violations of both descriptive and procedural invariance. In the TTO, violations of descriptive invariance have been observed in the form of labelling effects, for example (Gerard et al., 1993; Rabin et al., 1993; and Robinson and Bryan, 2001). Sackett and Torrance (1978) found that labelling a health state 'tuberculosis' led to significantly higher values than the same health state described as an 'unnamed contagious disease'.

Violations of procedural invariance have been widely reported in TTO valuations. For example, inconsistencies have been found when values are elicited through different search procedures (Lenert et al., 1998; Sumner and Nease Jr., 2001). Dolan et al. (1996) observed inconsistent values between a TTO that used props and diagrams and a 'no-props' variant. In another study, Stalmeier (2002) found that varying the endpoints of full health and death (0 and 1) had an effect on TTO values. Put simply, this technique – termed chaining – involves using anchor points of death and the best possible health state when full health is not a realistic outcome (e.g. in palliative care scenarios) (see Jansen et al., 1998). The value of the health state in question is chained onto the full-health/death scale through a subsequent TTO exercise. There is also evidence that the ordering of EQ5D dimensions may affect responses (Rand-Hendriksen and Augestad, 2012).

Of particular relevance to the research presented in Chapter 4 of this thesis, procedural invariance has been empirically violated in studies which have asked respondents to trade HRQoL for increased longevity (hereon termed a reverse TTO) instead of the standard approach of trading length of life for HRQoL. In these instances, inconsistent TTO values for the same health state depend on whether the standard or reverse procedure was used (Attema and Brouwer, 2013; Bleichrodt et al., 2003; Oliver and Wolff, 2014). Moreover, the magnitude of the differences in standard and reverse TTO values appear to be dependent on the time horizons under consideration.

An impact of the time horizon on TTO values has also been widely observed and thus CPTO has been shown not to hold (Attema and Brouwer, 2010; Kirsch and McGuire, 2000; Stalmeier et al., 2001; Stiggelbout et al., 1995). TTO values have been found to decrease at longer time horizons in some studies (e.g. Tsuchiya and Dolan, 2005) while the opposite pattern, increasing TTO values with increasing time horizon, has been observed in others (e.g. Stalmeier et al., 1997; Unic et al. 1998). These results suggest that life years are not fungible and that the trade-off between longevity and HRQoL may be influenced by a number of factors unaccounted for in the TTO model. Violations of CPTO raise key questions concerning the intertemporal nature of values and whether values elicited at one time horizon can appropriately be applied to others.

Thus, despite the relatively straightforward concept of a trade-off of longevity and HRQoL under conditions of certainty, the TTO's simple façade masks a significant level of complexity. There is a compelling need for valid health state preferences for application in CUA and to guide health policy decisions, and the TTO is widely used in its several variants for this purpose. Thus, it is important to not only acknowledge but direct greater attention towards understanding the circumstances in which TTO inconsistencies arise and thus the assumptions of the TTO model fail to hold. Behavioural economic concepts provide possible explanations for a number of observed inconsistencies in other contexts and the application of this (behavioural economic) knowledge in reference to TTO valuations is warranted. The following section presents the behavioural economic lens through which this thesis examines violations of the TTO's underlying assumptions and potential ways forward based on this framework.

1.4 Behavioural Economics Overview

Over the past several decades, studies of individual decision-making and rational choice in a variety of contexts have produced a mountain of evidence documenting 'irrational' behaviour; that is, behaviour that deviates from the predictions of standard economic theory. Notably, many of these deviations may be predicted on the basis of particular elements of the decision context (Tversky and Kahneman, 1974). To account for such descriptive departures from the normative model, research has increasingly turned to psychological-based explanations.

At the intersection of psychology and economics, the discipline of behavioural economics has put forward a number of explanations for observed decision anomalies. Whereas standard choice models depict how individuals *ought* to behave (a normative account), behavioural economics is focused on understanding how individuals *do* behave (a descriptive account). Behavioural economics questions the premise that individuals are able to express stable, context-independent preferences. Research has shown that decisions are often made using mental shortcuts and limited information and that context may influence decisions (e.g. responses to questions may be vulnerable to how information is presented) (Kahneman and Tversky, 1974). Simon (1972) credits the use of mental shortcuts (heuristics) that enable individuals to more easily execute decisions to bounded rationality. Bounded rationality refers to cognitive limitations which inhibit individuals from evaluating all possible information to inform an 'optimal' decision. As a result, individuals may select an outcome they deem to be *good enough* (termed *satisficing*) as opposed to an optimal one as would be predicted by standard economic theory. Although this type of expedited decision-making may be useful to the extent that individuals face too many decisions to allow for the processing of all relevant information, decision shortcuts and heuristics are problematic for the TTO and health state valuation more generally if elicited preferences fail to coincide with underlying preferences for the health state in question (Della Vigna, 2009).

Oliver (2013, p. 689) explains that rather than providing an overarching alternative theory, behavioural economics instead proposes a "set of observations that show that the cognitive processes that people employ when making decisions often systematically, and therefore seemingly deliberately, violate the set of assumptions and axioms that underlie the dominant neoclassical model." As such, behavioural economics encompasses a growing number of behavioural and cognitive anomalies that, while they can be categorized in a number of different ways, lack a unifying theory. In the absence of a parsimonious theory, behavioural economic concepts have largely been discussed on the basis of their relevance to the domain under consideration. The following section presents the behavioural economic principles most relevant to the TTO.

As a brief preamble, there are a number of key principles of behavioural economics and related judgment and decision-making literature that have emerged in the health state valuation literature. Concepts such as reference dependence and loss aversion, for example, have long been recognized as possible influences on health state values. Prospect theory, a prominent theory nested within the behavioural economics paradigm, encompasses reference dependence and loss aversion (along with probability weighting and diminishing sensitivity to increasing gains and losses). Prospect theory emerged as an empirically supported descriptive theory of choice between risky options and has been actively applied to health state values elicited under risk in the standard gamble to account for inconsistencies from expected utility theory (the normative theory which underpins the standard gamble) (Attema et al., 2013; Kahneman and Tversky, 1979; Oliver, 2003; van Osch and Stiggelbout, 2008; van Osch et al., 2004). The following section presents many of this theory's key elements in greater detail as they are also central to its riskless analogue, the reference dependence model. In comparison with the standard gamble, the literature on the application of behavioural economic concepts to the TTO is limited. More specifically, while there has been much ad hoc discussion as to how the TTO might be influenced by behavioural economic concepts, the applied research is much less.

1.41 Behavioural Economic Concepts in the Context of the TTO

The behavioural economic concepts that are readily applicable to the TTO, and are considered in this thesis, are the reference dependence model, how outcomes are evaluated within this model, the role of time preference, and how the preference is elicited or 'mapped' in terms of search procedure. Figure 1-3 presents a brief description of these behavioural influences, how the standard choice theory and TTO model assumptions and behavioural economic assumptions are expected to result in inconsistent values and the implications for the validity of TTO values. A final point of discussion is the potential impact of heuristic decision strategies on TTO values.

	Standard choice theory/TTO model assumption	Behavioural Economic interpretation	Implication for TTO
Reference dependence (Loss aversion, endowment effect, status quo bias, diminishing sensitivity)	Final endowment as value carrier Context-independent, complete preference orders	Values are a function of gains and losses from a reference point Overweighting of losses (loss aversion) Diminishing sensitivity	Loss aversion upwardly influences TTO values Endowment effect and status quo bias increase the resistance to making trade-offs (increasing TTO values) Violations of procedural and descriptive invariance
Time preference	Linear (TTO model) Discounted utility (standard choice theory; generalized TTO/QALY model)	Hyperbolic	Positive time preference: lower TTO values over time Negative time preference: higher TTO values over time Further important considerations: - maximum endurable time - adaptation Violations of CPTO
Search procedure	No effect	Search procedure effects: Scale compatibility Prominence effect	Both scale compatibility and prominence effect increase TTO values Violations of procedural invariance
Choice behaviour	Weighted additive Compensatory (losses in longevity compensated by gains in quality of life)	Limited attention Non-compensatory decision strategies and heuristic use	Lexicographic or other strategic trading Values inaccurate representations of HRQoL Ambiguous directional influence on TTO values

Figure 1-3: Comparison of standard economic theory (TTO model) assumptions, implications drawn from behavioural economics, and corresponding influences on TTO values

1.42 The Reference Dependent Model

Reference dependence: Standard economic theory assumes that value rests in the final state of endowment (i.e. preferences/health state values are reference *independent*). This assumption, however, has been widely refuted by the notion that outcomes are evaluated in relative terms instead of as absolutes (Markowitz, 1952; Tversky and Kahneman, 1991). Formally, the evaluation of one outcome in relation to a reference point is the fundamental premise of the reference dependence model, a riskless interpretation of

Kahneman and Tversky's Prospect Theory (1979, 1984; Tversky and Kahneman, 1991). Reference dependence along with loss aversion, diminishing sensitivity, and the status quo bias, each described in turn below, form the core of the reference dependence model.

To illustrate reference dependence, suppose A denotes the individual's current health state. If they are asked to value a worse health state, B, the reference dependent model posits that they will not value B in itself, but rather the value of B will be contingent on its location relative to A. Since health state B is worse than A, the individual will incur a loss in health status. On the other hand, suppose that the individual's current health state is C, which is worse than B. In this case, B is preferable to C, and therefore from C, B looks like a gain, whereas B constitutes a loss from their original reference point, A.

Loss aversion: Whether an outcome is considered a gain or a loss from the reference point is an important distinction due to loss aversion. Loss aversion implies that losses are weighted more heavily than gains. That is, there is an asymmetric valuation of outcomes causing a 'kink' in the value function at the reference point from which outcomes are assessed. As can be seen in Figure 1-4 the steeper slope for losses stemming from the reference point indicates an increased sensitivity versus commensurate sized gains.

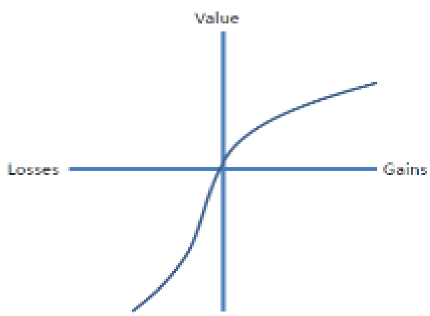


Figure 1-4: A hypothetical value function

In the TTO, loss aversion posits that the value the respondent assigns to a health state will be upwardly influenced. This is because losses will loom larger than gains and therefore the respondent will be reluctant to trade-off longevity for improved HRQoL. In the standard TTO model, the respondent is first presented with (Q_1, T_1) (seen graphically in Figure 1-5), which (under model assumptions) constitutes their reference point. The respondent is asked to trade-off a gain in health status (from Q_1 to full health) against a loss in longevity. In line with the explanation given by Bleichrodt (2002), in the absence of loss aversion, T_2 is the number of years provided by the respondent, whereby the gain in utility resulting from an increased health status is equal to the loss of utility in terms of life years. If we are to assume that the respondent is loss averse, they will weigh the loss in terms of time (T_1-T_2^1) more than the gain in HRQoL (Q_1-FH) and will therefore prefer the initial starting point (Q_1, T_1) over (FH, T_2^1). T_2^1 would need to increase to T_2^2 in order for the respondent to be indifferent between the two outcomes. As a result, loss aversion will have an upward influence on TTO values since T_2^2/T_1 will exceed T_2^1/T_1.

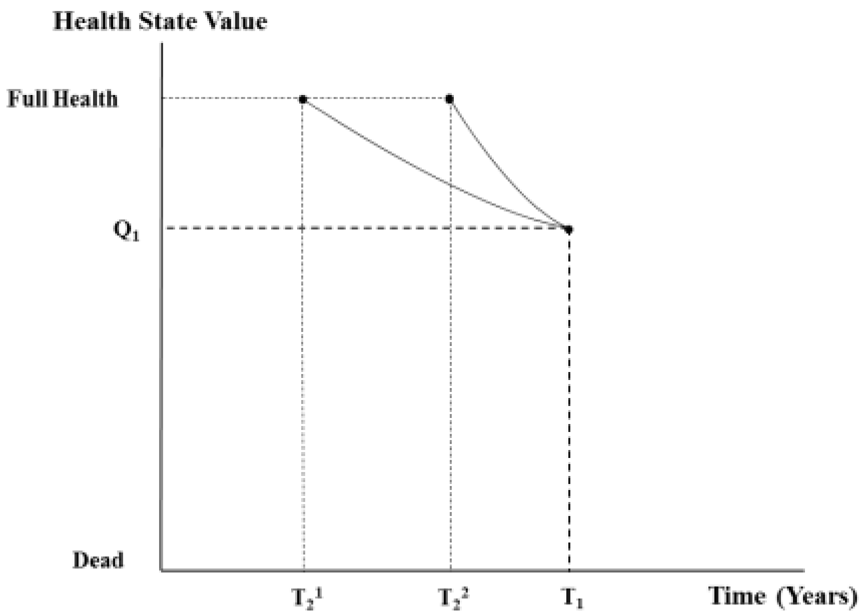


Figure 1-5: Loss aversion in the time trade-off (adapted from Bleichrodt, 2002)

Loss aversion can cause violations of procedural invariance and CPTO (Bleichrodt and Pinto, 2002). In regards to CPTO, greater loss aversion at shorter durations will render the respondent less willing to forgo duration for increased HRQoL (e.g. Bleichrodt et al., 2003). However, as the time horizon increases, longevity and HRQoL have been shown to become increasingly substitutable and thus the respondent becomes more willing to trade longevity for HRQoL, yielding relatively lower TTO values (Attema and Brouwer, 2013; Bleichrodt and Pinto, 2002; McNeil et al., 1981). Violations of procedural invariance have been attributed to loss aversion when comparisons are made between values elicited through a standard and a reverse TTO (i.e. how many years in the health state of interest the respondent would require in order to be indifferent between a shorter period in full health). A small number of studies have shown that systematically lower values are observed in the reverse TTO construct. Further, differences between standard and reverse constructs are especially apparent at short durations, where loss aversion has the greatest upward influence on standard TTO values (further explored in Chapter 2).

Status quo bias and endowment effect: Two phenomena closely related to loss aversion are the status quo bias and the endowment effect. The status quo bias implies that individuals prefer to maintain their current position. A similar concept, the endowment effect implies that individuals place additional value on what they own compared to what they do not yet own (even when an object, such as a coffee mug, is a randomly assigned, *instant* endowment) (Kahneman et al., 1991; Tversky and Kahneman, 1991). As a result, individuals are reluctant to forgo any of their current endowment, requiring more to give up (lose) a good than they are willing to give to acquire it (gain) (Kahneman and Tversky, 1979; Knetsch, 1992). Thus, the endowment effect can be seen as an integral component of loss aversion (conversely, loss aversion may be seen as an explanation for the endowment effect) (Knetsch, 1989). Importantly, the disproportionate weight allocated to what is part of one's endowment is not accounted for in standard economic theory. In the TTO, the status quo bias and endowment effect may manifest in the respondents' unwillingness to trade longevity for improved HRQoL.

Diminishing sensitivity: An important characteristic of the value function seen in Figure 3 is the flattening out of the function, implying diminishing sensitivity to increasing gains and losses. Diminishing sensitivity means that the further away the gain or loss occurs from the reference point, the smaller the marginal impact (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). Or, correspondingly, the incremental impact of a change in an outcome is smaller as the size of the outcome increases. In the TTO, this means that the incremental value of an additional life year is perceived as being progressively less as the time horizon increases, yielding lower values than at shorter time horizons. For example, with a (reference point) life expectancy of 10 years, the impact of a gain of 2 additional years from 10 to 12 years is greater than from a 2-year gain from 60 to 62 years.

1.43 Time-variant Preferences

Another important consideration when using the TTO is the respondent's time preference. Time preference refers to how the value assigned to an outcome changes depending on when the outcome occurs. The TTO model assumes that the proportion of time the respondent is willing to trade-off is stable across varying time horizons (hence CPTO), therefore assuming that a respondent does in fact have preferences towards the timing of ill health. Nonlinearity in TTO values, however, has been widely observed in the form of both positive and negative time preference rates. Individuals willing to trade-off an increasing proportion of time as the time horizon increases exhibit positive time preference rates. This implies that proximal years are more highly valued than future years (i.e. a preference to receive health benefits today rather than later) therefore TTO values will be biased in a downward direction. Conversely, an individual who has a negative time preference would be willing to trade-off smaller proportions of the time as the time horizon increases. In the TTO, a respondent with a negative time preference rate will be relatively more willing to forgo length of life for better quality at shorter durations resulting in lower values than those elicited from respondents with neutral or positive time preference rates (Dolan and Gudex, 1995; Unic et al., 1998).

Both adaptation and maximum endurable time in a given state illustrate specific cases of nonlinear time preferences in the TTO (Bleichrodt et al., 2003; Dolan and Stalmeier, 2003; Sutherland et al., 1982; Spencer, 2003). In such instances, the individual displays a non-monotonic preference (essentially, where the value of the health state neither constantly increases nor decreases with time) (Sutherland et al., 1982). Maximum endurable time (MET) implies that a point exists at which the individual does not wish to live any longer because the health state is so undesirable. On the other hand, if a respondent anticipates adaptation to a health state, they will be progressively less willing to make the trade-off as the time horizon increases. The effects of MET and adaptation on TTO values are summed up by Dolan and Stalmeier (2003) who explain that constant proportional trade-offs "may not be a very good representation of people's preferences in that the value of some less severe health states may increase over time and the value of some more severe states may decrease over time" (p. 447).

A generalized TTO (QALY) model accounts for constant (exponential) discount rates and has been applied in the literature by some researchers (Attema and Brouwer, 2010, 2012). A constant discount rate implies that the same discount rate is applied in each period such that no matter when they are asked, an individual feels the same about a particular intertemporal trade-off (Rabin, 1998). It is worth mentioning that hyperbolic models, where discount rates decrease as duration increases, have been shown to offer a better goodness of fit than exponential models when using monetary outcomes (Thaler, 1980). However, although there are important implications that follow from discount rates that are inconsistent over time (dynamic inconsistency), few studies have examined this in the context of health state valuation (Bleichrodt et al., 2014) and hyperbolic rates are not applied in practice (e.g. EQ5D) in economic evaluations.

1.44 Search Procedure Effects

A significant body of literature has shown that preferences elicited using choice and matching search procedures are inconsistent (Fischer and Hawkins, 1989; Lenert et al., 1998; Payne et al., 1992; Sumner and Nease Jr., 2001; Tversky et al., 1988). Briefly, it has been proposed that TTO values elicited through choice and matching differ due to the inconsistent weighting of attributes (i.e. longevity and HRQoL) between the two strategies. Most notably, scale compatibility and the prominence effect offer insight into inconsistencies. The prominence effect proposes that respondents have a tendency to select the outcome that is greater in the attribute they deem most important when responses are elicited through choice (Lichtenstein and Slovic, 1971; Slovic and Lichtenstein, 1968; Tversky et al., 1988). For instance, when eliciting TTO values through matching, the prominence effect posits that respondents will allocate a relatively greater amount of attention to years of life. Tversky et al. (1988) infer that matching may encourage quantitative decision-making processes and thus more weight to the attribute used as the ‘currency’ on which to match – years of life. Scale compatibility refers to the tendency to assign greater weight to an attribute that is consistent with the response scale in the task. In the TTO, respondents are asked to provide their answer in terms of duration; therefore, scale compatibility implies that more attention will be given to duration as opposed to health status. As a result, TTO values will be upwardly biased because the respondent will be less willing to trade-off life years for improvements in health status (Bleichrodt, 2002; Bleichrodt et al., 2003; Tversky et al., 1988). Other possible factors relating to the different values elicited through choice and matching are explored in the empirical studies in Chapters 2, 3, and 4.

1.45 Heuristics/Non-compensatory Decision Strategies

A final consideration is in regards to *how* – i.e. the type of decision strategy – the respondent arrives at their TTO valuation. Importantly this consideration carries different implications than the influences that have been discussed to this point which have assumed that the respondent trades length and HRQoL in an additive (i.e. compensatory) manner, such that a loss in longevity can be compensated for through a gain in health status.

Additive processing underlies the TTO model and requires that all information relating to each outcome – i.e. both HRQoL and longevity - be considered (Amaya-Amaya et al., 2002; Takemura, 2014).

The trade-off between HRQoL and longevity, however, is a cognitively challenging process (Stiggelbout and De Vogel-Voogt, 2008). Such trade-offs are likely to be unfamiliar and complex and therefore it is perhaps unsurprising that decisions are often made in a non-additive manner and subject to mental shortcuts (Payne et al., 1988; Payne and Bettman, 2004). Seminal work in behavioural economics demonstrated that due to information overload and bounded-rationality, individuals often rely on cognitive shortcuts or rules of thumb to reduce time and effort in making judgments and choices (Tversky and Kahneman, 1973; Tversky and Kahneman, 1974).

Two examples of non-compensatory decision-making strategies are elimination-by-aspects and lexicographic strategies. Elimination-by-aspects involves eliminating any outcomes that fall below a certain cut-off on the most important attribute. For example, a respondent who wishes to live until a certain age or until they are able to achieve a particular objective may be unwilling to select an outcome in which the number of years falls below a particular threshold. In such instances, preferences may be entirely functions of life expectancy rather than the target health state. This has led some critics to comment that "the TTO is a measure of reluctance to give up life rather than the severity of their health state" (Fowler et al., 1995, p. 198).

The compensatory nature of the TTO is undermined is if the respondent bases their trade on longevity-related goals (Loomes and McKenzie, 1989). Longevity-related goals encompass goals the respondent holds for which it is necessary to live a minimum number of years to achieve (e.g. a parent wishes to see their child graduate from university in 3 years and is therefore unwilling, regardless of the health state, to accept less than 3 years in full health in the TTO). An important distinction must be made between HRQoL and longevity-related goals in terms of their implications for the validity of health state values (Hazen, 2007). The use of goals relating to quality of life to guide trade-offs in the TTO may be seen as a valid assessment of the HRQoL of the health state in question (e.g. if the health state does not allow the respondent to participate in typical activities). On the other hand, longevity-related goals fail to account for how the health state might impact HRQoL thus jeopardizing the validity of the resulting values.

From a behavioural economic perspective, goal-based trade-offs can be interpreted as attribute substitution, which implies that individuals assess the target outcome (e.g. the health state) by substituting another, more readily accessible outcome in its place from which to base their decision (Kahneman, 2003). In the TTO, the respondent may substitute consideration of the severity of the health state with a longevity-dependent goal by which they can justify their trade. To this extent, Dixon et al. (2009, p. 14) comment “there may be reasons to believe that the willingness to trade length of life for quality of life is not simply a function of remaining life expectancy as is assumed in the existing policy.”

In lexicographic decision-making, attributes are allocated relative levels of importance and outcomes are analysed based on what the individual has deemed the most important attribute so that the outcome with the highest level on the most important attribute is selected (also termed attribute non-attendance) (Payne et al., 1993). In the TTO, lexicographic preferences can be observed, for example, when the respondent exclusively opts for the outcome with the greatest longevity (e.g. Chapman et al. 1999; Luce et al., 1997; O’Leary et al, 1995; Stiggelbout et al., 1995, 1996). In other words, no amount of increase on a given attribute (e.g. HRQoL) can compensate for any possible decrease on the other attribute (e.g. length of life).

An additional heuristic which has received attention in the TTO literature is that of proportional trade-offs (Dolan and Stalmeier, 2003; Stalmeier et al., 1997). In using a proportional heuristic, the respondent trades-off a constant proportion of the time horizon by, for example, doubling their stated number of years in full health when the time horizon doubles (and thereby feigning CPTO). Dolan and Stalmeier (2003) comment that respondents who implement a proportional heuristic to arrive at a response rely on this simple ‘rule of thumb’ without [adequate] consideration of the severity of the health state.

Given that non-compensatory decision strategies represent simplifications of the TTO task, their use may result in systematic errors or biases and thus result in inaccurate estimates of the HRQoL associated with a given health state. Despite this, the particular decision strategies adopted by respondents in the TTO have received little attention in the literature (e.g. Chapman et al., 1999; Dolan and Stalmeier, 2003; O’Leary et al., 1995). The study of non-compensatory decision strategies is much more prominent in the discrete choice experiment literature since these tasks require the consideration of trade-offs over larger sets of attributes within each outcome (Araña and León, 2009; Ryan et al., 2009). This is an important limitation of the current literature given the possible implications in terms of the validity of health state values elicited.

1.5 Summary: Where the TTO Currently Stands

For the TTO to be useful in CUA it must be assumed that the values which it elicits are accurate reflections of the HRQoL of a given health state. Essentially, the *goodness* of the policy that is chosen or cost-effectiveness decision that is made will be a reflection of the validity of the TTO values used in the analysis. Such validity depends on respondents being able to provide stable preferences for health states independent of the method by which they are elicited. However, the frequent empirically demonstrated violations of the TTO's underlying assumptions and the limited research directed at understanding the psychological processes underlying TTO valuations suggest that the values are both less valid and reliable than generally assumed and an unsteady ground from which to base economic evaluation and health resource allocation.

While there is a clear rationale for the TTO to be a recommended tool for health state valuation – decision-makers require some form of standardized method and data to inform their choices – the method and values elicited are contestable on both empirical and theoretical bases. It is therefore important to continue to seek ways to improve both the reliability and validity of the TTO; adopting the TTO, or any particular method, for reasons of expediency is of uncertain value if expressed preferences are poor representations of the HRQoL of the health state in question.

There have been numerous calls to attend to the TTO’s issues of validity and reliability. Mulhern et al. (2013), among others, have remarked that although the underlying assumptions of the TTO often fail to be met, the details of the causes and pervasiveness of such failures remain elusive. For example, Buckingham and Devlin (2009) commented that while “…the assumptions themselves and evidence of violations of them are discussed in the literature… the issues appear not to be widely appreciated by those using and applying TTO in economic evaluation.” Johnson (2009) also highlighted the need for methodologically-driven empirical research concerning the elicitation of health state values:

> “… the simplicity of the QALY as a universal health-care metric certainly has led to its broad acceptance, but it comes at the cost of several limiting assumptions. The significance of these assumptions are generally poorly understood, or simply ignored, among practitioners and policy-makers despite the assumptions’ repeated failure in careful tests of validity and reliability.” (Johnson, p. S38, 2009)

Few researchers have sought to describe the underlying structural integrity of the method. Instead, recent research endeavours pertaining to TTO methodology have largely followed other research agendas - such as reconciling TTO measurement for states worse than death, or trialling lead- and lag-time TTO procedures (e.g. Attema et al., 2013; Devlin et al., 2011), or have focused on the use of a single standardized version. Beresniak et al. (2015) remark that research aimed at addressing the methodological shortcomings and theoretical violations in the valuation of HRQoL has been overshadowed by the number of studies published that report applied cost/QALY analyses without questioning the underlying assumptions of the tools used to elicit the health state values. There has been considerable interest in understanding why the elicitation methods differ from one another (e.g. comparisons between VAS, standard gamble and TTO values) in terms of their susceptibility to loss aversion, time preference, and probability weighting (Bleichrodt, 2002; van Osch et al., 2004). Limited research, however, has been directed at testing the conditions in which violations of the assumptions occur *within* varying iterations of the TTO method.

The behavioural economic framework presented in this section is put forward with the aim of better understanding why the TTO has been plagued by inconsistencies. In interpreting the TTO from a behavioural economics perspective it is evident that a number of factors thought to be normatively inconsequential may indeed affect TTO values. Behavioural economics sheds light on a number of possible contributing factors to these inconsistencies: loss aversion, time preference, scale compatibility and the possibility that respondents rely on cognitive short-cuts or heuristics to arrive at their trade-off decisions. Despite these concepts being well established in other disciplines, we still understand remarkably little about their effects on health state valuation and specifically TTO values. Violations of procedural and descriptive invariance and CPTO undermine the validity of the TTO in terms of yielding values that are useful in CUA. For insights from judgment and decision-making research to remain on the perimeter of the discussion in health state valuation is a disservice to policy-makers who rely on these methods to guide societal decisions and to the individuals on whom these policies impact.

The following section outlines possible ways forward in light of the departures from the TTO assumptions described above and sets out the approach taken in this thesis.

1.6 Moving forward

Broadly, if the QALY model is to maintain, within reason, its current framework, two distinct approaches have been put forward.[xi] One possible strategy, and that seemingly favoured by health economists (and consistent with behavioural economic traditions), is termed here as a 'correction approach.' This approach seeks to address the impact of time preference and loss aversion on health state values through measurement and quantitative correction; i.e. transformation and fitting of data, in a sense 'correcting back to rationality'.[xii] While quantifying deviations from rational behaviour has been an ongoing motivation in behavioural economics, there are a number of challenges relating to how and when such corrections might be applied in the context of the TTO. A second approach involves understanding the influence of different TTO constructions on the values elicited without necessarily aiming to align values with normative predictions.

It should be noted that although there is some debate over the normative status of loss aversion in health state valuation[xiii], it has been suggested that if loss aversion causes inconsistencies in the tool, then it should be avoided or corrected (e.g. O'Connor, 1989). Given that the TTO model and the assumptions of standard economic theory do not accommodate loss aversion, it is therefore not considered in this thesis to be a component of an individual's underlying preference.

A key barrier to the correction of TTO values for loss aversion is the limited knowledge concerning how to estimate the magnitude of its influence with precision. The quantification of loss aversion has been examined in other disciplines using monetary outcomes where loss aversion coefficient (λ) of approximately two (meaning losses are weighted twice as heavily as commensurate sized gains) has been estimated (Abdellaoui et al., 2007, 2008; Booij and van de Kuilen, 2009; Gächter et al., 2007; Tversky and Kahneman, 1991 and 1992). In a health context, the only estimated coefficient is that of Attema et al. (2013) in the context of the standard gamble and no such quantitative estimation for the TTO could be found.

Attempting to quantify loss aversion is further complicated by the fact that there is ambiguity surrounding the location of the reference point (which is likely dependent on both respondent characteristics and question construction). Although it is often assumed that the fixed outcome in poor health is the reference point adopted by the respondent, there has only been one study that has explicitly examined this hypothesis (van Osch, 2007). Others have proposed that the reference point may be the respondent's expectation for longevity (which is external to the task, e.g. their subjective life expectancy) (Heinz et al., 2013; van Nooten et al., 2009). Moreover, there is evidence suggesting λ may be malleable, varying over different questions and frames and potentially dependent on the importance of the particular dimension (Bleichrodt and Pinto, 2002; Tversky et al., 1988; Tversky and Kahneman, 1991).

A number of methods aimed at measuring intertemporal preferences have been proposed and empirically examined in the TTO (e.g. Attema and Brouwer, 2010). While many of these approaches are analogous to those used in the context of monetary outcomes, a matter rarely touched upon in the health state valuation literature is that, given the nature of health and the value it carries over life stages, certain time periods may hold greater inherent meaning relative to others. In this regard, Oliver and Cookson (2010, p.645) comment:

> "… the simple application of a constant discount rate does not of course take into account the possibility that people have complex rates of time preference over life years, due to their placing particularly high value on some periods of their life, such as their life- defining years, their child-rearing years, and/or their retirement years, for example."

It is also notable that some studies (Attema and Brouwer, 2012a; Attema and Brouwer, 2012b) found that correcting for time preferences resulted in more severe violations of CPTO. Thus, the issue of time preference in the TTO may be less about fitting curves to data but rather a need to return to more conceptual considerations of how time itself is valued and used as a currency.

It is important to also consider that loss aversion and time preference often operate in tandem with other influences (e.g. scale compatibility). For instance, it has been proposed that at longer durations the upward influence of loss aversion and scale compatibility may be counterbalanced by a positive time preference rate, which exerts an opposing downward influence. Thus, to correct for one but not for other possible influences may result in even further inaccuracies (Bleichrodt, 2002). To illustrate, suppose that TTO values are adjusted for time preference (increasing values, assuming the respondent has a positive rate of time preference), and no adjustments are made for loss aversion or search procedure effects. The resulting values would be higher than in the absence of the correction (and therefore may be a poorer estimate of the HRQoL associated with the health state) (van Osch et al., 2004).

A final point is that a correction approach assumes that respondents have consistently engaged in compensatory decision-making strategies in the TTO tasks, i.e. they have considered all of the information presented and trade length and HRQoL in an additive manner such that a loss in longevity can be compensated for through a gain in health status. If the respondent relies on different decision strategies when completing TTOs which vary, for example, according to their time horizon, then correcting the respondents' values for what appears to be loss aversion or the effects of time preference may be unjustified (e.g. where the respondent adopts a compensatory decision strategy when the TTO uses a relatively long time horizon and a non-compensatory strategy when a relatively shorter time horizon is used, adjustments of both values for time preference would seem inappropriate). Essentially, corrections may be estimated and curves fit to data (i.e. TTO values) but if the underlying decision processes are not the same (i.e. if respondents are basing their trade-offs on something other than the undesirability of the health state) then this correction would be potentially misleading.

In summary, there is a lack of current knowledge on how we might apply corrections for behavioural economic influences and, more generally, a number of outstanding conceptual challenges. Further, potentially important factors such as search procedure effects and heuristic decision strategies are not amenable to quantitative adjustments. Thus, the position adopted in this thesis is that we should instead adopt a more qualitative approach to understanding inconsistencies in TTO values. Note that the pursuit of a more descriptive approach in understanding the influence of behavioural economic concepts on TTO values does not imply that their correction is inherently wrong. Rather, it is necessary to further understand the underlying nuances of the TTO so as to determine whether correction is appropriate – and if so, how it might best be undertaken – or whether it is misguided.

The direction of the research undertaken in this thesis is one of gaining insight into the structural integrity of the TTO, as has been expressed repeatedly in the literature - "more research is needed that develops and tests criteria to assist in determining the preferable procedure for the performance of TTOs" (Attema and Brouwer, 2012, p. 498). This thesis has the goal of highlighting an alternative approach to testing and quantifying deviations from predictions of standard economic theory, by focusing on understanding the descriptive sources of inconsistencies which emerge in TTO valuations. A behavioural framework is used to analyse inconsistencies, and, more generally, to provide insight into how the accuracy of stated preference methods both in health state valuation and elsewhere might be improved.

Payne et al. (1999) refer to such an approach as establishing a "building code" whereby a better understanding of how different constructions of choice problems can affect choice behaviour and responses is acquired (see also Payne et al., 1988, 1993). In order to do so, the TTO task is presented in several different ways to gain insight as to how different constructions of the question affect values. Although seemingly structurally simple, the TTO has many ‘moveable parts’, meaning that different variants (using different search procedures and time horizons, for example) may be differentially susceptible to deviations from the assumptions (Nord, 1991). This intra-technique variability has proven to be detrimental to the comparability of TTO values in the literature; it is possible, however, that the pluriformity of constructions available within the standard TTO model is advantageous if it can assist in finding a particular protocol (or set of protocols) that is preferable, yielding valid estimates of HRQoL. Moreover, this approach will attend to potential underlying conceptual problems such as under what circumstances is the task able to adequately capture HRQoL as opposed to measuring some other construct (e.g. how long is needed to achieve a particular goal).

1.7 Conclusions and Structure of the Thesis

Methods used in health state valuation have evolved since the concept of HRQoL measurement was introduced over 40 years ago. The need for robust methods and quantification of HRQoL for both clinical and policy purposes will remain as limited resources face further demand. As a relied upon tool for eliciting health state values, the TTO is applied in ways and in circumstances where variable results may have profound consequences.

The assumptions of invariance and CPTO underpin the validity of the TTO as a health state elicitation tool. These assumptions are “normatively indispensable” in that they ensure the value is not a reflection of how it is elicited (i.e. an artifact of the elicitation process) but rather of the health state of interest (Kahneman and Tversky, 1986; Mellers et al., 1995). Thus, the degree to which the TTO in its various constructions yields values that are consistent with these assumptions is a measure of their validity.

The objective of this thesis is to examine the consistency of the TTO in terms of its assumptions, with the particular goal of gaining insight into the nature and extent to which the behavioural economic factors discussed above compromise such consistency. This thesis aims to build on the existing literature, which shows that violations of the TTO’s underlying assumptions are extensive, by searching for consistency using hypotheses based on behavioural economic phenomena (including, for example, loss aversion and heuristics). More precisely, the goal is to examine the degree to which the TTO assumptions of invariance and CPTO are observed. While such internal consistency alone is not sufficient, inconsistency is a clear threat to the validity of the TTO, and a preferred TTO construction can be defined as one which demonstrates superior internal consistency

(Oliver, 2004). For example, CPTO may be confounded by the use of proportional heuristics as observed by Dolan and Stalmeier (2003) and it is therefore desirable to corroborate validity by other means. This is explored and further discussed in the individual studies.

In evaluating the internal consistency of the TTO, a number of subsidiary questions are addressed such as under what conditions: “Does the value of quality of life depend on duration?” (CPTO); “Does the value of quality of life depend on the search procedure/response mode?” (procedural invariance); and “Does the value of quality of life depend on how the TTO is framed?” (procedural and descriptive invariance). Ultimately, the findings will contribute to the broader conceptual discussion of the questions: “Can sacrificed life years be used as a stable currency?” (CPTO) and “Does a hypothetical trade between longevity and quality of life reveal accurate preferences for given health states?”

The empirical studies presented in this thesis underpin current research that looks to modify the basic TTO structure in attempts, for example, to accommodate more accurate measurement of health states worse than death. As we continue to use the TTO and develop variants based on its conventional form, the question of which method(s) most accurately capture HRQoL estimates without confounding influences should be considered of great importance.

The study presented in Chapter 2 tests descriptive invariance, focusing on a comparison between a standard TTO and an age-framed TTO. Different specifications of the TTO may invite different psychological processes and decision-making strategies, some of which may be more susceptible to the effects of construct-irrelevant factors than others. It is critical that the TTO asks *the right question* – i.e. that ambiguity in how the respondents interpret what they are being asked is minimized and the considerations (e.g. capabilities, pain, prior experience, etc.) factoring into their decision are logical (in a sense, predictable) given the decision context (Prades, 1997; Wittrup-Jensen and Pedersen, 2008). This study explores the effects of an age-framed TTO, whereby the time horizon is expressed as the respondent’s age at death (calculated by adding their current age to the time horizon in question). CPTO was assessed in both standard and age-framed TTO, supported by qualitative data, to gain insight into the circumstances under which CPTO (as well as procedural and descriptive invariance) are violated.

The notion of presenting information to respondents in an easily accessible manner draws on the principle of concreteness from psychological research. The empirical work in Chapter 3 examined the effect of framing the TTO using a life-expectancy time horizon.

A 10-year time horizon, implemented, for instance, using EQ5D methods, has been criticized on the grounds that many respondents find this remaining length of life artificially short. The study in Chapter 3 thus aimed to understand how values differed when measured at both a 10-year time horizon and a life-expectancy time horizon. A life-expectancy TTO offers an additional operationalization of the TTO which has received limited attention in the literature and is therefore deserving of further investigation. It was hypothesized that by using a long, more realistic time horizon, i.e. the respondent's life expectancy, respondents would defer less often to cognitive shortcuts to arrive at their valuations and therefore the underlying rationale for their decision might result in values which are of greater validity.

The study reported in Chapter 4 investigates whether procedural invariance is observed between standard and reverse TTOs. While the consistency of values between these two variants has been evaluated in a handful of other reports, their methods vary substantially and it is thus difficult to draw any general conclusions. Specifically, the study in Chapter 4 aimed to determine whether procedural invariance between standard and reverse TTO values is mediated by the search procedure by which the values are elicited. The comparison of search procedures enabled a second test of procedural invariance to be undertaken. The study had a further objective of carrying out a test of the sensitivity of the search procedures' respective abilities to capture differences in HRQoL across health states of different severities. The effects of scale compatibility and loss aversion in a matching search procedure would predict that the longevity attribute is over-weighted in this context relative to choice, and thus it was hypothesized that this would undermine the procedure's ability to adequately capture changes on the quality of life dimension.

In the concluding chapter of the thesis, the implications of the three empirical studies are considered in relation to the need for stated preference methods to inform health care resource allocation and the larger context of the potential means of determining the allocation strategy given the existing health economic landscape in the UK.

References

Abdellaoui, M., Bleichrodt, H., & l'Haridon, O. (2008). A tractable method to measure utility and loss aversion under prospect theory. *Journal of Risk and Uncertainty, 36*(3), 245-266.

Abdellaoui, M., Bleichrodt, H., & Paraschiv, C. (2007). Loss aversion under prospect theory: A parameter-free measurement. *Management Science, 53*(10), 1659-1674.

Amaya-Amaya, M., Ryan, M., & San Miguel, F. (April, 2002). *Do individuals adopt non-compensatory decision strategies. Preliminary evidence from health care.* Paper presented at the Discrete Choice/Conjoint Workshop, Institute of Public Health – Health Economics, University of Southern Denmark: Odense.

Araña, J. E., & León, C. J. (2009). Understanding the use of non-compensatory decision rules in discrete choice experiments: the role of emotions. *Ecological Economics, 68*(8), 2316-2326.

Arrow, K., Solow, R., Portney, P., Leamer, E., Radner, R., & Schuman, H. (1993). Report of the NOAA Panel on Contingent Valuation *Federal Register* (Vol. 58, pp. 4601-4614).

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *The American Economic Review*, 941-973.

Arrow, K. J., Cropper, M. L., Eads, G. C., Hahn, R. W., Lave, L. B., Noll, R. G., . . . Smith, K. (1996). Benefit-cost analysis in environmental, health, and safety regulation. *Washington, DC: American Enterprise Institute*, 1-17.

Attema, A., & Brouwer, B. (2010). On the (not so) constant proportional trade-off in TTO. *Quality of Life Research, 19*, 489-497.

Attema, A. E., Brouwer, W. B., & l'Haridon, O. (2013). Prospect theory in the health domain: A quantitative assessment. *Journal of Health Economics, 32*(6), 1057-1065.

Attema, A. E., & Brouwer, W. B. F. (2012). The way that you do it? An elaborate test of procedural invariance of TTO, using a choice-based design. *The European Journal of Health Economics, 13*(4), 491-500. doi: 10.1007/s10198-011-0318-y

Attema, A. E., Edelaar-Peeters, Y., Versteegh, M. M., & Stolk, E. A. (2013). Time trade-off: one methodology, different methods. *The European Journal of Health Economics, 14*(Suppl 1), 53-64. doi: 10.1007/s10198-013-0508-x

Augustovski, F., Irazola, V., Velazquez, A., Gibbons, L., & Craig, B. (2009). Argentine valuation of the EQ-5D health states. *Value Health, 12*(4), 587–596.

Augustovski, F., Rey-Ares, L., Irazola, V., Oppe, M., & Devlin, N. J. (2013). Lead versus lag-time trade-off variants: does it make any difference? *Eur J Health Econ, 14 Suppl 1*, S25-31. doi: 10.1007/s10198-013-0505-0

Baron, J., & Ubel, P. A. (2002). Types of inconsistency in health-state utility judgments. *Organizational Behavior and Human Decision Processes, 89*(2), 1100-1118.

Bennett, J., Nease, R. F., Jr., & Sumner, W., 2nd. (2002). Evidence for the rapid construction of preference during utility assessments. *Proc AMIA Symp*, 41-45.

Bentham, J. (1780). *An Introduction to the Principles of Morals and Legislation* (Vol. 45): Dover Publications.

Beresniak, A., Medina-Lara, A., Auray, J. P., De Wever, A., Praet, J. C., Tarricone, R., . . . Duru, G. (2015). Validation of the underlying assumptions of the quality-adjusted life-years outcome: results from the ECHOUTCOME European project. *Pharmacoeconomics, 33*(1), 61-69. doi: 10.1007/s40273-014-0216-0

Bleichrodt, H. (2002). A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Economics, 11*, 447-456.

Bleichrodt, H., Gao, Y., & Rohde, K. I. M. (2014). *Decreasing Impatience for Health and Money*. Erasmus School of Economics. Rotterdam, The Netherlands.

Bleichrodt, H., & Johannesson, M. (1997). Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. *Journal of Health Economics, 16*(2), 155-175.

Bleichrodt, H., Luis Pinto, J., & Abellan-Perpinan, J. M. (2003). A consistency test of the time trade-off. *Journal of Health Economics, 22*, 1037-1052.

Bleichrodt, H., & Pinto, J. L. (2002). Loss Aversion and Scale Compatibility in Two-Attribute Trade-Offs. *Journal of Mathematical Psychology, 46*, 315–337.

Booij, A. S., & Van de Kuilen, G. (2009). A parameter-free analysis of the utility of money for the general population under prospect theory. *Journal of Economic Psychology, 30*(4), 651-666.

Bossert, W. (1991). On intra-and interpersonal utility comparisons. *Social Choice and Welfare, 8*(3), 207-219.

Bowling, A. (2001). *Measuring disease: a review of disease-specific quality of life measurement scales*. (2nd ed.). Buckingham, England: Open University Press.

Boye, K. S., Matza, L. S., Feeny, D. H., Johnston, J. A., Bowman, L., & Jordan, J. B.

(2014). Challenges to time trade-off utility assessment methods: when should you consider alternative approaches? *Expert Rev Pharmacoecon Outcomes Res, 14*(3), 437-450. doi: 10.1586/14737167.2014.912562

Brazier, J., Ratcliffe, J., Salomon, J., & Tsuchiya, A. (2007). *Measuring and Valuing Health Benefits for Economic Evaluation*: OUP Oxford.

Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *J Health Econ, 21*(2), 271-292.

Brazier, J., & Rowen, D. (2011). NICE DSU technical support document 11: alternative to EQ-5D for generating health state utility values. Report by the Decision Support Unit.

Briggs, A. H., & O'Brien, B. J. (2001). The death of cost-minimization analysis? *Health Econ, 10*(2), 179-184. doi: 10.1002/hec.584

Brock, D., & Wikler, D. (2006). Ethical issues in resource allocation, research, and New Product Development. In D. Jamison, J. Breman, A. Measham, G. Alleyne, M. Claeson, D. Evans, P. Jha, A. Mills & P. Musgrove (Eds.), *Disease Control priorities in developing countries* (pp. 259-270). New York: Oxford University Press and The World Bank.

Brock, D. W. (2003). Ethical issues in the use of cost-effectiveness analysis for the prioritization of health care resources. *WHO guide to cost-effectiveness analysis*, 289.

Brouwer, W. B., & Koopmanschap, M. A. (2000). On the economic foundations of CEA. Ladies and gentlemen, take your positions! *J Health Econ, 19*(4), 439-459.

Buckingham, K., Devlin, N., & Tabberer, M. (2004). A theoretical framework for TTO valuations and a taxonomy of TTO approaches: results from a pilot study (Vol. Report No. 04/07). London, UK: Department of Economics, City University London.

Buckingham, K. J., & Devlin, N. J. (2009). A note on the nature of utility in time and health and implications for cost utility analysis. *Soc Sci Med, 68*(2), 362-367. doi: 10.1016/j.socscimed.2008.09.048

Cabases, J., Gaminde, I., Ugalde, J. M., & Pozo, F. (2000). *Social elicitation of EQ-5D health state preferences through Person Trade-Off (PTO).* Paper presented at the 17th Plenary Meeting of the EuroQoL Group, Pamplona, Spain.

Camerer, C. (2005). Three cheers--psychological, theoretical, empirical--for loss aversion. *Journal of Marketing Research*, 129-133.

Chapman, G. B., Elstein, A. S., Kuzel, T. M., Nadler, R. B., Sharifi, R., & Bennett, C. L. (1999). A multi-attribute model of prostate cancer patients' preferences for health states. *Quality of Life Research, 8*(3), 171-180.

Chevalier, J., & de Pouvourville, G. (2011). Valuing EQ-5D using time trade-off in France. *Eur J Health Econ, 14*(1), 57-66.

Claxton. K, Martin, S., Soares, M., Rice, N., Spackman, E., Hinde, S., & al., e. (2013). *Methods for the estimation of the NICE cost effectiveness threshold*. York: University of York.

Coast, J., Flynn, T. N., Natarajan, L., Sproston, K., Lewis, J., Louviere, J. J., & Peters, T. J. (2008). Valuing the ICECAP capability index for older people. *Soc Sci Med, 67*(5), 874-882. doi: 10.1016/j.socscimed.2008.05.015

Committee, Pharmaceutical Benefits Advisory. (June 2013). *Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee.* Available at http://www.pbac.pbs.gov.au.

Cookson, R., & Culyer, A. (2010). Measuring overall population health: the use and abuse of QALYs. In A. Killoran (Ed.), *Evidence-Based Public Health: Effectiveness and Efficiency, p. 148*. Oxford University Press.

Coons, S., & Kaplan, R. (2005). Cost-utility analysis. In J. Bootman, R. Townsend & W. McGhan (Eds.), *Principles of Pharmacoeconomics* (3rd ed., pp. 117–147). Cincinnati, OH: Harvey Whitney Books.

Cracknell, R. (2010). ‘The ageing population – the UK’s ageing population has considerable consequences for public services’ *Key issues for the new Parliament 2010*: House of Commons Library Research Archive.

Culyer, A. J. (1992). The morality of efficiency in health care—some uncomfortable implications. *Health Economics, 1*(1), 7-18.

Culyer, A. J. (2006). The bogus conflict between efficiency and vertical equity. *Health Econ, 15*(11), 1155-1158. doi: 10.1002/hec.1158

Dakin, H., Devlin, N., Feng, Y., Rice, N., O'Neill, P., & Parkin, D. (2014). The influence of cost-effectiveness and other factors on NICE decisions. *Health Economics*. doi: 10.1002/hec.3086

Daniels, N. (2001). Justice, health, and healthcare. *Am J Bioeth, 1*(2), 2-16. doi: 10.1162/152651601300168834

De Martino, B., Camerer, C. F., & Adolphs, R. (2010). Amygdala damage eliminates monetary loss aversion. *Proceedings of the National Academy of Sciences, 107*(8), 3788-3792.

De Wit, G. A., Busschbach, J., & De Charro, F. (2000). Sensitivity and perspective in the valuation of health status: whose values count? *Health Economics* (9), 109–126.

DellaVigna, S. (2009). Psychology and Economics: Evidence from the Field. *Journal of Economic Literature, 47*(2), 315-372. doi: 10.1257/jel.47.2.315

Devlin, N., Buckingham, K., Shah, K., Tsuchiya, A., Tilling, C., Wilkinson, G., & van Hout, B. (2013). A comparison of alternative variants of the lead and lag time TTO. *Health Econ, 22*(5), 517-532. doi: 10.1002/hec.2819

Devlin, N. J., Tsuchiya, A., Buckingham, K., & Tilling, C. (2011). A uniform time trade off method for states better and worse than dead: feasibility study of the 'lead time' approach. *Health Econ, 20*(3), 348-361. doi: 10.1002/hec.1596

Dixon, S., Longworth, L., & Wailoo, A. (2009). Assessing technologies at the end of life: A review of empirical evidence. . School of Health and Related Research, University of Sheffield: NICE Decision Support Unit ScHARR.

Dolan, P. (1998). Valuing health-related quality of life. *Pharmacoeconomics*, 15(2), 119-127.

Dolan, P. (2000). The measurement of health-related quality of life for use in resource allocation decisions in health care. *Handbook of Health Economics, 1*, 1723-1760.

Dolan, P., Gudex, C., Kind, P., & Williams, A. (1996). Valuing health states: a comparison of methods. *Journal of Health Economics, 15*, 209-231.

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

Dolan, P., & Roberts, J. (2002). To what extent can we explain time trade-off values from other information about respondents? *Social Science & Medicine, 54*(6), 919-929.

Dolan, P., & Stalmeier, P. (2003). The validity of time trade-off values in calculating QALYs: constant proportional time trade-off versus the proportional heuristic. *Journal of Health Economics, 22*(3), 445-458.

Dormont, B., Grignon, M., & Huber, H. (2006). Health expenditure growth: reassessing the threat of ageing. *Health Economics, 15*(9), 947-963.

Dormont, B., & Huber, H. (2006). Causes of Health Expenditure Growth: The Predominance of Changes in Medical Practices over Population Ageing. *Annales d'Économie et de Statistique* (83/84), 187-217. doi: 10.2307/20079168

Drummond, F. (2005). *Methods for the Economic Evaluation of Health Care Programmes*: Oxford University Press.

Drummond, M., Torrance, G., & Mason, J. (1993). Cost-effectiveness league tables: more harm than good? *Social Science & Medicine, 37*(1), 33-40.

Drummond, M. F. (2005). *Methods for the Economic Evaluation of Health Care Programmes*: Oxford University Press.

Drummond, M. F., Torrance, G. W., & Mason, J. C. (1993). Cost-effectiveness league tables: More harm than good? *Social Science and Medicine, 37*(1), 33-40.

National Institute for Clinical Excellence (April 2013). Guide to the methods of technology appraisal 2013. https://www.nice.org.uk/process/pmg9/chapter/foreword

Fischer, G. W., & Hawkins, S. A. (1993). Strategy compatibility, scale compatibility, and the prominence effect. *Journal of Experimental Psychology: Human Perception and Performance, 19*(3), 580.

Fowler, F. J., Cleary, P. D., Massagli, M. P., Weissman, J., & Epstein, A. (1995). The role of reluctance to give up life in the measurement of the values of health states. *Medical Decision Making, 15*(3), 195-200.

Freeman, A. M. (1993). Nonuse values in natural resource damage assessment. *Valuing natural assets: the economics of natural resource damage assessment*, 264-303.

Freeman III, A. M. (1979). Benefits of environmental improvement: theory and practice.

Frisch, D., & Clemen, R. T. (1994). Beyond expected utility: rethinking behavioral decision research. *Psychological Bulletin, 116*(1), 46.

Furlong, W. J., Feeny, D. H., Torrance, G. W., & Barr, R. D. (2001). The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Ann Med, 33*(5), 375-384.

Gächter, S., Johnson, E. J., & Herrmann, A. (2007). Individual-level loss aversion in riskless and risky choices. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1010597

Garber, A. M., & Phelps, C. E. (1997). Economic foundations of cost-effectiveness analysis. *Journal of Health Economics, 16*(1), 1-31.

Gerard, K., Dobson, M., & Hall, J. (1993). Framing and labelling effects in health descriptions: quality adjusted life years for treatment of breast cancer. *Journal of Clinical Epidemiology, 46*(1), 77-84.

Gigerenzer, G. (2008). Why heuristics work. *Perspectives on psychological science, 3*(1), 20-29.

Green, C. (2001). On the societal value of health care: what do we know about the person trade-off technique? *Health Econ, 10*(3), 233-243. doi: 10.1002/hec.581

Gregory, R., Flynn, J., Johnson, S. M., Satterfield, T. A., Slovic, P., & Wagner, R. (1997). Decision-Pathway Surveys: A Tool for Resource Managers. *Land Economics, 73*(2), 240-254. doi: 10.2307/3147285

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy, 80*(2), 223-255. doi: 10.2307/1830580

Güth, W. (2008). (Non) behavioral economics: A programmatic assessment. *Zeitschrift für Psychologie/Journal of Psychology, 216*(4), 244-253.

Gyrd- Hansen, D., Kjær, T., & Nielsen, J. (2012). Scope insensitivity in contingent valuation studies of health care services: should we ask twice? *Health Economics, 21*(2), 101-112.

Haubrich, D., & Wolff, J. (2006). Economism and its Limits. In M. Moran, M. Rein & R. Goodin (Eds.), *The Oxford Handbook of Public Policy*. Oxford: Oxford University Press.

Hausman, D. M. (2010). Hedonism and welfare economics. *Economics and Philosophy, 26*(3), 321.

Hazen, G. (2007). Adding Extrinsic Goals to the Quality-Adjusted Life Year Model. *Decision Analysis, 4*(1), 3-16. doi: doi:10.1287/deca.1060.0082

Heintz, E., Krol, M., & Levin, L.-Å. (2013). The Impact of Patients' Subjective Life Expectancy on Time Tradeoff Valuations. *Medical Decision Making, 33*(2), 261-270.

Henshall, C., & Schuller, T. (2013). Health technology assessment, value-based decision making, and innovation. *International Journal of Technology Assessment in Health Care, 29*(04), 353-359.

Hurley, J. (2000). An overview of the normative economics of the health sector. *Handbook of Health Economics, 1*, 55-118.

Insinga, R. P., & Fryback, D. G. (2003). Understanding differences between self-ratings and population ratings for health in the EuroQOL. *Quality of Life Research, 12*(6), 611-619.

Jansen, S. J., Stiggelbout, A. M., Wakker, P. P., Vliet Vlieland, T. P., Leer, J. W., Nooy, M. A., & Kievit, J. (1998). Patients' utilities for cancer treatments: a study of the chained procedure for the standard gamble and time tradeoff. *Med Decis Making, 18*(4), 391-399.

Janssen, M.F., Birnie, E., Haagsma, J.A., Bonsel, G.J. (2008). Comparing the standard EQ-5D three-level system with a five-level version. *Value Health*, 11: 275–284.

10.1111/j.1524-4733.2007.00230

Janssen, B. M. F., Oppe, M., Versteegh, M. M., & Stolk, E. A. (2013). Introducing the composite time trade-off: a test of feasibility and face validity. *The European Journal of Health Economics, 14 (Suppl 1),* 5-13. doi: 10.1007/s10198-013-0503-2

Johnson, F. R. (2009). Editorial: Moving the QALY forward or just stuck in traffic? *Value Health, 12 Suppl 1*, S38-39. doi: 10.1111/j.1524-4733.2009.00521.x

Johnson, F. R., & Bingham, M. F. (2001). Evaluating the validity of stated-preference estimates of health values. *Swiss Journal of Economics and Statistics (SJES), 137*(I), 49-63.

Kahneman, D. (2009) A different approach to the QALY. *Value Health*, 12 (Suppl. 1): S16–7.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 1449-1475.

Kahneman, D., Knetsch, J., & Thaler, R. (1991). The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives, 5*, 193-206.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy*, 1325-1348.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47*(2), 263-292.

Kahneman, D., & Tversky, A. (1984). Choices, Values, and Frames. *American Psychologist, 39*(4), 341-350.

Kirsch, J., & McGuire, A. (2000). Establishing health state valuations for disease specific states: an example from heart disease. *Health Economics, 9*, 149-158.

Knetsch, Jack (1989), “The Endowment Effect and Evidence of Nonreversible Indifference Curves,” American Economic Review, 79 (December), 1277–84.

Knetsch, J. L. (1992). Preferences and nonreversibility of indifference curves. *Journal of Economic Behavior & Organization, 17*(1), 131-139.

Kopec, J. A., & Willison, K. D. (2003). A comparative review of four preference-weighted measures of health-related quality of life. *J Clin Epidemiol, 56*(4), 317-325.

Lafortune, G., van Gool, K., Biondi, N., & al., e. (2013). Health at a glance 2013: OECD indicators. Organization for Economic Cooperation and Development. Paris.

Lamers, L., McDonnell, J., Stalmeier, P., Krabbe, P., & Busschbach, J. (2006). The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Economics, 15*(10), 1121–1132.

Lenert, L. A., Cher, D. J., Goldstein, M. K., Bergen, M. R., & Garber, A. (1998). The effect of search procedures on utility elicitations. *Medical Decision Making, 18*(1), 76-83.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology, 89*(1), 46-55.

Liu, G. G., Wu, H., Li, M., Gao, C., & Luo, N. (2014). Chinese Time Trade-Off Values for EQ-5D Health States. *Value in Health, 17*(5), 597-604. doi: http://dx.doi.org/10.1016/j.jval.2014.05.007

Loomes, G., & McKenzie, L. (1989). The use of QALYs in health care decision making. *Soc Sci Med, 28*(4), 299-308.

Loomes, G., & Sugden, R. (1983). A Rationale for Preference Reversal. *The American Economic Review, 73*(3), 428-432. doi: 10.2307/1808127

Luce, M. F., Bettman, J. R., & Payne, J. W. (1997). Choice processing in emotionally difficult decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(2), 384.

Markowitz, H. (1952). The utility of wealth. *The Journal of Political Economy*, 151-158.

McDonough, C. M., & Tosteson, A. N. (2007). Measuring preferences for cost-utility analysis: how choice of method may influence decision-making. *Pharmacoeconomics, 25*(2), 93-106.

McKie, J., Kuhse, H., Richardson, J., & Singer, P. (1996). Allocating healthcare by QALYs: the relevance of age. *Cambridge Quarterly of Healthcare Ethics, 5*, 534-545.

McNeil, B., Weichselbaum, R., & Pauker, S. (1981). Tradeoffs between quality and quantity of life in laryngeal cancer. *New England Journal of Medicine, 305*, 982–987.

Mehrez, A., & Gafni, A. (1991). The Healthy-years Equivalents How to Measure Them Using the Standard Gamble Approach. *Medical Decision Making, 11*(2), 140-146.

Mellers, B., Weber, E., Ordonez, L., & Cooke, A. (1995). Utility invariance despite labile preferences. In J. Busemeyer, R. Hastie & D. Medin (Eds.), *The Psychology of Learning and Motivation* (Vol. 32, pp. 221-246). Waltham, MA: Academic Press.

Menzel, P., Gold, M. R., Nord, E., Pinto- Prades, J. L., Richardson, J., & Ubel, P. (1999). Toward a Broader View of Values in Cost- Effectiveness Analysis of Health.

Hastings Center Report, 29(3), 7-15.

Mishan, E. J. (1971). Evaluation of life and limb: a theoretical approach. *The Journal of Political Economy*, 687-705.

Gold, M. R., Patrick, D. L., Torrance, G. W., Fryback, D. G., Hadorn, D. C., Kamlet, M. S., Daniels, N., & Weinstein, M. C. (1996). Identifying and valuing outcomes. In Gold, M. R., Siegel, J. E. Russell, L. B., & Weinstein, M. C. (Eds.), *Cost-Effectiveness in Health and Medicine.* (pp. 82-134). New York: Oxford University Press.

Mulhern, B., Longworth, L., Brazier, J., Rowen, D., Bansback, N., Devlin, N., & Tsuchiya, A. (2013). Binary Choice Health State Valuation and Mode of Administration: Head-to-Head Comparison of Online and CAPI. *Value in Health, 16*(1), 104-113. doi: 10.1016/j.jval.2012.09.001

Norwegian Medicines Agency (NOMA). (2012). *Guidelines on how to conduct pharmacoeconomic analyses*. Available at: http://www.ispor.org/peguidelines/source/Norwegian_guidelines2012.pdf

National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. 2013. Available at: http://www.nice.org.uk/media/D45/1E/GuideToMethodsTechnologyAppraisal2013.pdf

O'Connor, A. M. (1989). Effects of framing and level of probability on patients' preferences for cancer chemotherapy. *Journal of Clinical Epidemiology, 42*(2), 119-126.

O'Leary, J. F., Fairclough, D. L., Jankowski, M. K., & Weeks, J. C. (1995). Comparison of time-tradeoff utilities and rating scale values of cancer patients and their relatives: evidence for a possible plateau relationship. *Med Decis Making, 15*(2), 132-137.

Oliver, A. (2003). The internal consistency of the standard gamble: tests after adjustment for prospect theory. *Journal of Health Economics, 22*, 659-674.

Oliver, A. (2004). Testing the internal consistency of the standard gamble in ‘success’ and ‘failure’ frames. *Social Science & Medicine, 58*(11), 2219-2229.

Oliver, A. (2013). Testing procedural invariance in the context of health. *Health Econ, 22*(3), 272-288. doi: 10.1002/hec.2796

Oliver, A. (2013). From Nudging to Budging: Using Behavioural Economics to Inform Public Sector Policy. *Journal of Social Policy, 42*(04), 685-700. doi:10.1017/S0047279413000299

Oliver, A., & Cookson, R. (2010). Analysing risk attitudes to time. *Health Economics, 19*(6), 644-655. doi: 10.1002/hec.1499

Oliver, A., & Wolff, J. (2014). Are people consistent when trading time for health? *Econ Hum Biol, 15*, 41-46. doi: 10.1016/j.ehb.2014.05.001

Olsen, J. A., Kidholm, K., Donaldson, C., & Shackley, P. (2004). Willingness to pay for public health care: a comparison of two approaches. *Health Policy, 70*(2), 217-228. doi: 10.1016/j.healthpol.2004.03.005

Oppe, M., Devlin, N. J., van Hout, B., Krabbe, P. F., & de Charro, F. (2014). A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value Health, 17*(4), 445-453. doi: 10.1016/j.jval.2014.04.002

Payne, J. W., & Bettman, J. R. (2004). Walking with the scarecrow: The information-processing approach to decision research. *Blackwell Handbook of Judgment and Decision Making*, 110-132.

Payne, J. W., Bettman, J. R., Coupey, E., & Johnson, E. J. (1992). A constructive process view of decision making: Multiple strategies in judgment and choice. *Acta Psychologica, 80*(1), 107-141.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 534.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*: Cambridge University Press.

Payne, J. W., Bettman, J. R., Schkade, D. A., Schwarz, N., & Gregory, R. (1999). Measuring constructed preferences: Towards a building code *Elicitation of Preferences* (pp. 243-275): Springer.

Post, P. N., Stiggelbout, A. M., & Wakker, P. P. (2001). The utility of health states after stroke a systematic review of the literature. *Stroke, 32*(6), 1425-1429.

Prades, J-L. P. (1997). Is the person trade-off a valid method for allocating health care resources? *Health Econ, 6*, 71-81.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American economic review*, 1281-1302.

Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature*, 11-46.

Rand-Hendriksen, K., & Augestad, L. A. (2012). Time trade-off and ranking exercises are sensitive to different dimensions of EQ-5D health states. *Value Health, 15*(5), 777-782. doi: 10.1016/j.jval.2012.04.002

Ratcliffe, J., Brazier, J., Tsuchiya, A., Symonds, T., & Brown, M. (2009). Using DCE and ranking data to estimate cardinal values for health states for deriving a preference-based single index from the sexual quality of life questionnaire. *Health Economics, 18*(11), 1261-1276.

Rawlings, D., & Cuyler, A. (2004). National Institute for Clinical Excellence and its value judgments. *British Journal of Medicine*, *329*, 224-227.

Revicki, D. A., & Kaplan, R. M. (1993). Relationship between psychometric and utility-based approaches to the measurement of health-related quality of life. *Quality of Life Research, 2*(6), 477-487.

Richardson, J. (1994). Cost utility analysis: what should be measured? *Soc Sci Med, 39*(1), 7-21.

Robinson, L., & Hammitt, J. (2011). Behavioral Economics and the Conduct of Benefit-Cost Analysis: Towards Principles and Standards. *Journal of Benefit-Cost Analysis, 2*(2).

Robinson, S., & Bryan, S. (2001). *Naming and framing': an investigation of the effect of disease labels on health state valuations*. Paper presented at the Health Economics Study Group Meeting, University of Oxford.

Ryan, M., Watson, V., & Entwistle, V. (2009). Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. *Health Economics, 18*(3), 321-336. doi: 10.1002/hec.1369

Sackett, D., & Torrance, G. (1978). The utility of different health states as perceived by the general public. *Journal of Chronic Diseases, 31*(11), 697-704.

Samuelsen, C., Augestad, L., Stavem, K., Kristiansen, I., & Rand-Hendriksen, K. (2012). *Anchoring effects in the lead-time time trade-off*. Paper presented at the Proceedings of the 29th EuroQol Plenary Meeting, The Doelen Concert and Congress Hall, Rotterdam, the Netherlands.

Seymour, J., McNamee, P., Scott, A., & Tinelli, M. (2010). Shedding new light onto the ceiling and floor? A quantile regression approach to compare EQ-5D and SF-6D responses. *Health Econ, 19*(6), 683-696. doi: 10.1002/hec.1505

Shah, K. K., Cookson, R., Culyer, A. J., & Littlejohns, P. (2013). NICE's social value judgements about equity in health and health care. *Health Economics, Policy and Law, 8*(02), 145-165. doi: doi:10.1017/S1744133112000096

Shiell, A., & Gold, L. (2003). If the price is right: vagueness and values clarification in contingent valuation. *Health Econ, 12*(11), 909-919. doi: 10.1002/hec.777

Shiell, A., Seymour, J., Hawe, P., & Cameron, S. (2000). Are preferences over health states complete? *Health Econ, 9*(1), 47-55.

Simon, H. A. (1972). Theories of bounded rationality. *Decision and Organization, 1*(1), 161-176.

Sintonen, H. (2001). The 15D instrument of health-related quality of life: properties and applications. *Ann Med, 33*(5), 328-336.

Slovic, P., & Lichtenstein, S. (1968). Relative importance of probabilities and payoffs in risk taking. *Journal of Experimental Psychology, 78*(3p2), 1.

Smith, R. D., & Richardson, J. (2005). Can we estimate the 'social' value of a QALY? Four core issues to resolve. *Health Policy, 74*(1), 77-84. doi: 10.1016/j.healthpol.2004.12.009

Sokol-Hessner, P., Hsu, M., Curley, N. G., Delgado, M. R., Camerer, C. F., & Phelps, E. A. (2009). Thinking like a trader selectively reduces individuals' loss aversion. *Proceedings of the National Academy of Sciences, 106*(13), 5035-5040.

Spencer, A. (2003). The TTO method and procedural invariance. *Health Economics*, 655-688.

Stalmeier, P. F. (2002). Discrepancies between chained and classic utilities induced by anchoring with occasional adjustments. *Med Decis Making, 22*(1), 53-64.

Stalmeier, P. F., Chapman, G. B., de Boer, A. G., & van Lanschot, J. J. (2001). A fallacy of the multiplicative QALY model for low-quality weights in students and patients judging hypothetical health states. *International Journal of Technology Assessment in Health Care, 17*(04), 488-496.

Stalmeier, P. F., Wakker, P. P., & Bezembinder, T. G. (1997). Preference reversals: Violations of unidimensional procedure invariance. *Journal of Experimental Psychology: Human Perception and Performance, 23*(4), 1196-1205.

Stamuli, E. (2011). Health outcomes in economic evaluation: who should value health? *British Medical Bulletin, 97*(1), 197-210.

Stiggelbout, A., Kiebert, G., Kievit, J., Leer, J., Habbema, J., & De Haes, J. (1995). The 'utility' of the time trade-off method in cancer patients: feasibility and proportional trade-off. *Journal of Clinical Epidemiology, 48*(10), 1207-1214.

Stiggelbout, A. M., & De Vogel-Voogt, E. (2008). Health State Utilities: A Framework for Studying the Gap Between the Imagined and the Real. *Value in Health, 11*(1), 76-87. doi: 10.1111/j.1524-4733.2007.00216.x

Stiggelbout, A. M., Eijkemans, M. J., Kiebert, G. M., Kievit, J., Leer, J. W., & De Haes, H.

J. (1996). The 'utility' of the visual analog scale in medical decision making and technology assessment. Is it an alternative to the time trade-off? *Int J Technol Assess Health Care, 12*(2), 291-298.

Stiggelbout, A. M., Kiebert, G. M., Kievit, J., Leer, J. W., Stoter, G., & de Haes, J. C. (1994). Utility assessment in cancer patients: adjustment of time tradeoff scores for the utility of life years and comparison with standard gamble scores. *Med Decis Making, 14*(1), 82-90.

Streiner, D., & Norman, G. (1989). *Health Measurement Scales: A Practical Guide to Their Development and Use.* Oxford: Oxford University Press.

Sumner, W., & Nease, R. F. (2001). Choice-matching preference reversals in health outcome assessments. *Medical Decision Making, 21*(3), 208-218.

Sutherland, H., Llewellyn-Thomas, H., Boyd, N., & Till, J. (1982). Attitudes towards survival. The concept of "maximum endurable time". *Medical Decision Making, 2,* 299-309.

Takemura, K. (2014). *Behavioral Decision Theory: Psychological and Mathematical Descriptions of Human Choice Behavior*: Springer, Japan.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization, 1*(1), 39-60.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization, 1*(1), 39-60. doi: http://dx.doi.org/10.1016/0167-2681(80)90051-7

Torrance, G. W. (2006). Utility measurement in healthcare: the things I never got to. *Pharmacoeconomics, 24*(11), 1069-1078.

Torrance, G. W., Feeny, D., & Furlong, W. (2001). Visual Analog Scales Do They Have a Role in the Measurement of Preferences for Health States? *Medical Decision Making, 21*(4), 329-334.

Tsuchiya, A. (1999). Age-related preferences and age weighting health benefits. *Social Science & Medicine, 48*(2), 267-276.

Tsuchiya, A., & Dolan, P. (2005). The QALY model and individual preferences for health states and health profiles over time: a systematic review of the literature. *Med Decision Making, 25*(4), 460-467. doi: 10.1177/0272989x05276854

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207-232.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases.

Science, 185(4157), 1124-1131. doi: 10.1126/science.185.4157.1124

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, S251-S278.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: a reference-dependent model. *The Quarterly Journal of Economics*, 1039-1061.

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

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

Ubel, P. A., Loewenstein, G., & Jepson, C. (2003). Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Quality of Life Research, 12*(6), 599-607.

Unic, I., Stalmeier, P. F., Verhoef, L. C., & van Daal, W. A. (1998). Assessment of the time-tradeoff values for prophylactic mastectomy of women with a suspected genetic predisposition to breast cancer. *Med Decis Making, 18*(3), 268-277.

van Nooten, F., Koolman, X., & Brouwer, B. (2009). The influence of subjective life expectancy on health state valuations using a 10 year TTO. *Health Economics, 18*, 549-558.

van Osch, S. M. (2007). *The Construction of Health State Utilities*. (PhD), Leiden University Medical Center (LUMC), Leiden University.

van Osch, S. M., & Stiggelbout, A. M. (2008). The construction of standard gamble utilities. *Health Economics, 17*(1), 31-40. doi: 10.1002/hec.1235

van Osch, S. M., Wakker, P. P., van den Hout, W. B., & Stiggelbout, A. M. (2004). Correcting biases in standard gamble and time tradeoff utilities. *Medical Decision Making, 24*(5), 511-517. doi: 10.1177/0272989x04268955

Von Neumann, J., & Morgenstern, O. (1994 (2007)). *Theory of games and economic behavior*: Princeton University Press.

Wagstaff, A. (1991). QALYs and the equity-efficiency trade-off. *J Health Econ, 10*(1), 21-41.

Whitty, J., Rundle-Thiele, S., & Scuffham, P. (2008). Insights into public preferences or pharmaceutical funding. *International Journal of Pharmaceutical and Healthcare Marketing, 2*(3), 216-234.

Wisloff, T., Hagen, G., Hamidi, V., Movik, E., Klemp, M., & Olsen, J. A. (2014). Estimating QALY gains in applied studies: a review of cost-utility analyses published in 2010. *Pharmacoeconomics, 32*(4), 367-375. doi: 10.1007/s40273-014-0136-z

Wittrup-Jensen, K. U., Pedersen, K. M., & Sundhedsøkonomi, S. U. F. f. (2008). *An Empirical Assessment of the Person Trade-off: Valuation of Health, Framing Effects, and Estimation of Weights for Fairness*: Syddansk Universitet.

End of Chapter 1 Notes

i. Given that some may question the use of the term market in the context of government subsidized health care, it is important to clarify that in this instance, market is used in reference to the provision of health care services regardless of how payment mechanisms are structured.

ii. Two additional approaches to assigning a monetary value to health outcomes have also been widely discussed: the human capital approach and the revealed preference approach) (see Drummond et al., 2005). The human capital approach involves estimating the present value (i.e. discounted sum) of the individual's future labour income. This approach therefore equates the value of a QALY with the economic productivity of a healthy individual over the course of one year. The revealed preference approach infers the value of a QALY from observing individual's actual behavior in real-world situations. As discussed in Section 4, the revealed preference approach faces a number of challenges in the valuation of health.

iii. An alternative approach to establishing a cost per QALY threshold involves the use of QALY league tables. QALY league tables rank interventions in ascending order from the lowest cost per QALY to the greatest cost per QALY. In theory, league tables offer a useful exercise aimed at informing which interventions should be prioritized. Interventions with the lowest incremental cost-effectiveness ratios (ICERs), situated at the top of the table, are implemented first (Smith and Richardson, 2005). An ICER threshold can be inferred based on the last intervention approved by policy-makers within a particular league table. The use of league tables as a means of informing resource allocation has been hindered, however, by a number of methodological challenges, including difficulties in constructing tables with sufficient comparability across study methods (e.g. costing perspectives: what types of costs and what perspective were included into the ICER numerator) (Drummond et al., 1993).

iv. EuroQol's EQ-5D older version, the EQ-5D 3L, describes health according to five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression, where each dimension has 3 levels (no problem, some or moderate problem and extreme problem). After about 2 decades of using the 3L version, there has been a shift towards a 5L version where descriptions of health are based on 5 levels (no problem, slight problem, moderate problem, severe problem and extreme problem). The development and adoption of the 5L version results from findings which have shown that the 3L version may be insensitive to minor but important differences in health as well as the observation of ceiling effects (meaning respondents are unwilling to trade any longevity for health) (Janssen et al., 2008).

v. A number of possible factors have been proposed to account for differences between patient and general population values including inadequate descriptions of the health states (i.e. the health state is insufficiently communicated to the general population resulting in a lack of scope – whereby there is a discrepancy between health state descriptions and the health states themselves which is caused by descriptions which are too sparse and therefore inadequately capture what it's like to experience a given heath state) and adaptation (Insinga and Fryback, 2003; Ubel et al., 2003). Further, it has been suggested that non-patients are susceptible to a focusing illusion (i.e. they may disproportionately focus on the negative aspects of the health state while ignoring the unaffected aspects) or may focus on the transition from their current state of health to the health state, thus valuing the shock and fear associated with the change rather than the health state (Stamuli, 2011; Ubel et al., 2003).

vi. Another technique, the person trade-off (PTO), asks respondents to indicate the number of one type of outcome would be equivalent to a given number of outcomes of another kind (e.g. how many outcomes of A (y) are equal to *x* number of outcomes of B). The number of individuals in health state A is varied until a point of equivalence is found between the two groups. Following that, it can be posited that health state B is x/y times as desirable as health state A. Through such a process, health states can be related to each other on a common value scale (Green, 2001).

The PTO task may be constructed in a number of different ways. Outcomes, for example, may involve either restoring individuals to full health or a previous health state, or, preventing death (see Tsuchiya, 1999, for various different constructions of the PTO task) (Cabases et al., 2000).

In contrast to the TTO and standard gamble, the outcome measure in the PTO is concerned with the welfare of others as opposed to the individual themselves. It has been proposed that rather than aiming to measure utility or value (as in the case of the standard gamble and TTO), the PTO assesses the *societal value* of a particular health state (Baron and Ubel, 2002).

Further, ordinal-based methods such as discrete choice experiments (e.g. Ratcliffe et al., 2009) and best-worst scaling may also be used to gauge preferences for health states (e.g. Coast et al., 2008) although unlike the standard gamble and time trade-off they do not directly provide cardinal values (although values may be modeled).

vii. The standard gamble often yields higher values than the TTO for two reasons (e.g. van Osch et al., 2004; Post et al., 2001). First, individuals tend to be risk averse, which means that they will hesitate to accept the gamble, driving values upward (since the probability of the gamble's success will be higher than in the instance where the individual is risk neutral). Second, individuals are often assumed to possess positive rates of time preference and are thus more willing to forgo life-years at the end of a time period, such as the time horizon in the TTO, lowering TTO values relative to standard gamble values.

viii. Health state valuation is grounded in von Neumann-Morgenstern's Expected Utility Theory of rational decision-making under uncertainty, for which the standard gamble has been interpreted as a relatively direct implementation given its incorporation of a risk element (von Neumann and Morgenstern, 1944). There is a vast literature dedicated towards whether health state values elicited using methods other than the standard gamble can be regarded as utilities (e.g. Richardson, 1994; Dolan, 1988). This debate is not considered to be essential or of direct relevance to the work in this thesis. Nonetheless, in recognition of this debate, health state values elicited using the TTO are referred to as values as opposed to utilities in this thesis.

ix. While the trade-off between these attributes is the central premise to the TTO model, it has been challenged by many, especially in the case of mild health states, where intuitively, one might not expect to even consider forgoing length of life. This fundamental consideration of the TTO is revisited a number of time throughout the thesis.

x. In recent years, the EuroQoL group has conducted preliminary investigations of lead and lag–time TTOs, whereby a period of full health is tacked on to the beginning (lead) or end (lag) of the trade-off period, respectively. A main intention in using lead or lag-time trade-offs was to accommodate the measurement of health states better and worse than death using the same method. However, these more recent iterations of the TTO are not without their own problems and face a number of challenges when used for health state valuation.

xi. It is worth noting; however, that a third approach exists which entails significant modifications of the QALY model, or, as discussed by Kahneman (2009), abandoning the QALY model altogether and using alternative means to prioritize healthcare interventions, for example, relying on the judgment of a small group of experts. While there are some proponents for this option, it is not a central focus in this thesis.

xii. Critics of this approach propose that rather than attempting to understand the underlying psychological processes of decisions – i.e. how and why choices are made – a correction approach instead attempts to preserve the standard expected utility framework by incorporating parameters and transformation. In such discussions, this approach is often referred to as a neoclassical *repair program* (Gigerenzer, 2008; Güth, 2008).

xiii. Camerer (2005) proposes that loss aversion is a product of fear which affects neurobiological processes (Sokol-Hessner et al., 2009; De Martino et al., 2010). An alternative perspective in terms of consumer sovereignty is advanced by Kahneman et al. (1990), who propose that where loss aversion is considered to be part of the underlying preference (as Camerer, p. 131, remarks “a genuine expression of preference” as opposed to an error in judgment), then it should not be corrected.

Chapter 2. Stating the Obvious? Tests of Descriptive and Procedural Invariance in the TTO

2.1 Abstract

In its standard form, the TTO provides the number of years for which the respondent can expect to live, the time horizon, which is followed by immediate death. It is therefore implicit that the respondent considers his or her maximum age at death as part of their TTO calculation. This study sought to investigate presenting the TTO question in a logically equivalent manner, simplifying this element of the TTO – i.e. by stating TTO time horizons in terms of the respondent's age at death. It was hypothesized that when the time horizon assumes an age-frame presentation, respondents would be loss averse to forgoing length of life for improved health.

Using a 2-by-2 design, a further test of procedural invariance was included with respondents completing either an age-frame or a standard TTO using a bisection or matching search procedure. The study was conducted via self-complete web-based survey, a method which is becoming increasingly popular among researchers focused on questions relating to health state valuation methodology.

In contrast to the initial hypothesis, when a bisection search procedure was used, the age-frame had a significant impact in the opposite direction than predicted by loss aversion: age-framed values were *lower* than standard values (age frame TTO10 values were significantly lower than standard frame values (0.53 vs 0.73, p=0.014; TTO30s not significantly different). When a matching procedure was used, no significant differences emerged between the two frames at either the 10-year or 30-year durations. One quarter of respondents gave same number of years in TTO10 and TTO30 which suggests respondents may use cognitive shortcuts in matching to arrive at their valuations. Results of the test of procedural invariance between the two search procedures indicated that this key TTO assumption was violated in both age-framed and standard TTO variants (bisection TTO values were consistently lower than matching at both 10- and 30-year time horizons, standard and age frame (all p<0.001); matching-bisection differences were 0.17 and 0.15 for the standard 10-year TTO and 30-year TTO, respectively, and 0.37 and 0.22 for age-frame 10- and 30-year TTOs.

The qualitative results indicated that only about a third of respondents in each group weighed length and quality of life in their decision processes, and revealed that

reasons other than the health state were often factored in. Relatively few studies have made attempts to ‘get behind the numbers’ and this finding underscores the importance of endeavouring to do so in future research.

While the results of the study did not support the initial hypothesis, they shed light on several alternative questions that should be examined in greater detail. In particular, it may be that an age-frame TTO yielded lower TTO values than a standard frame because the respondent was able to allocate more attention to the health state versus calculating their age based on the time horizon (as might be anticipated in the standard TTO). It seems plausible that by presenting the TTO in an age-frame, the respondent doesn’t need to allocate as much attention to the length of life dimension since they don’t need to calculate how old they will be at the end of the time horizon. As a result, the respondent may be better able to attend to the health state in question. For matching, where no differences were observed between age-frame and standard conditions, it is suspected, as posited by other authors (Attema and Brouwer, 2012), that respondents engage in an anchor and adjustment process whereby the matching process requires them to focus on the length of life provided and adjust for quality accordingly from this duration.

More generally, this study highlights the importance of how values are elicited, paying particular attention to respondent interpretation of the task and with many respondents indicating greater ease with longer time horizons. Also, the study showed that conducting the TTO using an internet based questionnaire was feasible, based on a small number of exclusions, a relatively large sample size, and similar completion times.

Future research should look to refute or support these findings, particularly since the greatest inconsistencies were observed using a bisection search procedure (commonly used in the literature) and when using a 10-year time horizon, a duration often used in the literature and recommended in standardized frameworks. Age-framed and standard TTO in face-to-face interviews should also be compared in future studies since this may offer greater insight into the feasibility of age-framed TTO relative to the standard TTO, and the rationales underlying respondent trade-offs in each. Additionally, qualitative data (to complement quantitative data) has provided valuable insight in terms of understanding the feasibility of web-based surveys and investigators should seek to capture this data in future studies wherever possible.

2.2 Background

In health care, as in all public sectors, resources are limited and therefore priorities must be set. Currently, across several national health care systems, health care resources are prioritized with the help of a particular type of cost-effectiveness analysis - cost-utility analysis - that is based, in part, on a composite measure of disease burden - the quality adjusted life year (QALY). The QALY combines health-related quality of life (hereon HRQoL) and length of life considerations and enables comparisons between otherwise disparate outcomes associated with different illnesses and conditions.

The time trade-off (hereon TTO) is a commonly used method for eliciting health state values (i.e. quantifications of HRQoL) for implementation into QALY calculations. In its standard form, the TTO asks respondents to state the number of years they would be willing to give up from a fixed time period (the time horizon) in a deteriorated health state in order to live in full health. For example, a respondent may be asked how many years they would be willing to forgo to live in full health as opposed to living with migraines for 10 years (the time horizon is determined by the researcher and may vary depending on the health state or the purpose of the study). The number of years the respondent is willing to trade from the time horizon (in this instance, 10 years) to achieve better health reflects their perceived value (i.e. the (un)desirability) of the health state in question. So, supposing the respondent is willing to forgo three years of their life in order to live without migraines, they would be assigning a TTO value of 0.70 ((10-3)/10) to the migraine health state.

The degree to which respondents understand and internalize the TTO (i.e. taking into account both the time horizon and the severity of the health state in question) is generally taken for granted despite evidence that indicates this is a topic deserving of greater debate. One possibility is that respondents interpret the trade-off, of say 10 years, to occur at the end of their lives (van Nooten et al., 2014). In such an instance, a young respondent may anticipate living many more years prior to this time horizon. Several other phenomena have the potential to distort TTO values. For example, decreasing marginal utility (implying each subsequent year of life is valued progressively less), neglecting the fact that death follows the period spent in the inferior health state, and expectations for a low HRQoL at older ages would each imply that years spent in relatively worse health (i.e. in the health state presented in the TTO) might be traded-off with greater ease (Brouwer and van Exel, 2005; van Nooten and Brouwer, 2004).

Further alternatives are that, rather than explicitly trading-off longevity for HRQoL, the respondent relies on a rule of thumb (e.g. proportional heuristics) or bases their decision of the number of years required to achieve a particular goal (therefore providing a valuation unrelated to the HRQoL of the health state) (Dolan and Stalmeier, 2003; Stalmeier et al., 1997). The latter possibility, formulating trade-offs according to goal achievement, was discussed by Simon (1957) as the aspiration level required by the decision-makers – i.e. "the value of a goal variable that must be reached or surpassed by a satisfactory decision alternative" (Gigerenzer, 2002). McFarlane et al. (2007), for instance, found among a sample of hemodialysis patients that respondents expressed that they wished to live until at least their next pivotal life event (e.g. a family wedding) and were therefore unwilling to trade any time before that event. The use of heuristics or basing trade-offs on attaining goals (unrelated to the health state) compromises the validity of TTO values. As an example, suppose the respondent based their trade on a particular life event and the timing of this event was moved forward by a year: such a change would affect the amount of time the respondent would be willing to forgo in the TTO. This timing effect would decrease the TTO score, yet is unlikely to have an impact on the HRQoL associated with the health state of interest.

Collins (2003) comments that the objective of implementing a standard question format (as in the 10-year EQ5D descriptive framework) is to ensure that any observed differences are non-artefactual (i.e. do not stem from deviations in how the preferences were elicited but rather due to true differences in what is being measured). This implies that all respondents interpret the question in the same way and that the question itself provides the respondent with all of the necessary information they require to arrive at their decision. The validity of TTO values in economic analysis is contingent on the consistent interpretation of the TTO question across different respondents. Heterogeneity across respondents' interpretations of the TTO is therefore an important consideration.

Only recently, van Nooten et al. (2014) set out to better understand how respondents interpret the time horizon in the TTO; specifically, whether respondents who were asked to state their age at death as implied by the time horizon were relatively less willing to trade longevity for HRQoL. They found that drawing respondents' attention to their age at death in the TTO, as implied by the time horizon, appears to induce greater loss aversion. Interestingly, when respondents were not prompted explicitly to think about their own age at the end of the fixed time period (10 years) only about half of respondents indicated in a follow-up question that they had done so. The results of van Nooten et al. in combination with qualitative data and quantitative inconsistencies found in the TTO literature indicate more work needs to be done to understand the cognitive processes which that underlie TTO valuations.

Whether or not respondents rely on simplification strategies or mental shortcuts to arrive at their responses when they are faced with a challenging decision has been of increasing, although as yet limited, interest in health state valuation. However, despite the potential for various different (mis)interpretations of the TTO task, relatively little is known about how respondents perceive the time horizon and thus what reasons underlie their trade-offs (Chapman et al., 1999). Whether the time horizon is interpreted as intended is an essential question given the importance of eliciting robust values for input into economic analysis.

2.3 The Study

The study is grounded in part in Slovic's (1972) principle of concreteness which implies that respondents have a tendency to use the information presented to them *as is*. To illustrate, the typical phrasing of TTO tasks – whereby the respondent is presented with a time horizon in terms of the number of years they have left to live, say "10 years" – may not sufficiently prompt respondents to internalize the implication of the time horizon on their age at death. As a result, they may adopt one of the perspectives mentioned above (such as imagining these years as occuring at the end of their subjective life expectancy (SLE) or failing to realize immediate death follows the time horizon).

A question of particular interest is therefore how might the TTO be presented in a more intuitive manner. Constructing health state valuation tasks (and, more generally, any sort of preference elicitation task) so that they can be used with facility is important to obtaining valid, accurate responses. The aim of the current study was to assess whether TTO values differ systematically when the TTO question is framed in a manner which explicitly states the respondent's age at death as the time horizon, in comparison to the standard presentation. Such a comparison provides a test of descriptive invariance, a fundamental assumption of the TTO and decision theory more generally, which implies that logically equivalent presentations of the same question (i.e. equivalent ways of describing outcomes) should yield consistent health state values.

In framing the TTO in a manner that provides the respondent with their age at death, it was thought that the respondent would be relieved of the cognitive effort associated with calculating the impact of the time horizon on their age at death. It was also anticipated that by expressing the time horizon as an age, thus assigning greater meaning to the time horizon, the respondent might be less inclined to use noncompensatory decision strategies (which refer to tactics used by the respondent to arrive at their valuation without trading-off length of life and HRQoL; e.g. lexicographic processing). A further, secondary, hypothesis was that by age-framing the TTO time horizon, respondents would provide *higher* values because their attention would be drawn to their age at death in the task, inducing greater loss aversion since the frame would highlight the conflict between the respondent's expectation for longevity outside of the task and the limitation imposed on this expectation by the time horizon.

Only two studies could be identified in the literature that have implemented the respondent's age at death as the time horizon in the task. Attema and Brouwer (2012) remarked that respondents found that questions used to elicit individual discount rates were more readily imagined when they were framed in terms of what the respondents' ages would be given the time horizon, although their main TTO elicitations did not adopt an age-frame. The other study, by Buitinga et al. (2012), used an actuarial life expectancy (i.e. the respondents estimated life expectancy based on age and gender norms) as the time horizon. They did not, however, include a comparator group using the standard presentation of the TTO time horizon and therefore the effect of the age-frame is unknown.

2.4 Methods

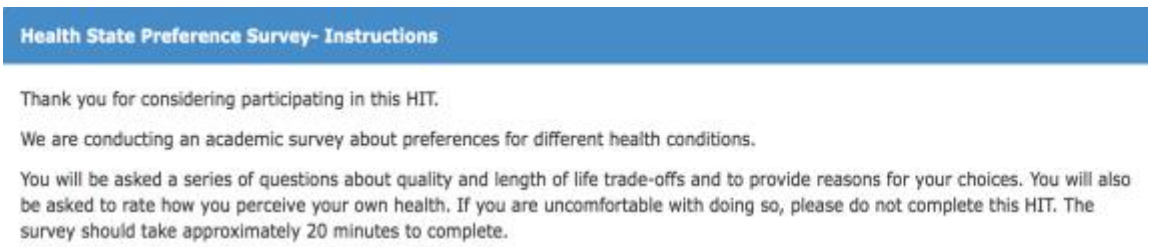
The study was hosted on the Qualtrics Survey Platform. A computer-based platform like Qualtrics enables a number of psychological parameters to be tested (e.g. the response times per window (i.e. per question) and total time per questionnaire). Using online survey techniques is an increasingly common strategy in eliciting preferences in the context of health state valuation: Craig et al., 2014; Richardson et al., 2014; van de Wetering et al., 2016. Augestad and Rand-Hendriksen (2013), Rencz et al. (2015), van Nooten et al. (2009, 2014, 2015), Augestad et al. (2016) and van Nooten et al. (2016) have used self-completed web-based surveys to carry out TTO valuations.

While there have been concerns that lower response rates and sampling biases can result from different modes of administration (Bowling, 2005; Oppe et al., 2014), there is evidence that suggests that web-based TTOs (Norman et al., 2010) and surveys on a variety of other topics (Amir et al., 2012; Horton et al., 2011; Simons and Chabris, 2012; Sprouse, 2011; Mason and Suri, 2011) yield results comparable to those in face-to-face interviews. Moreover, advantages of this approach, including increased sample sizes and the potential for greater diversity in respondent groups, may outweigh possible disadvantages, making this a useful alternative to smaller-scale samples often comprised of postgraduate students.

2.41 The Sample

491 respondents were recruited through Amazon.com’s Mechanical Turk platform (hereon MTurk, www.MTurk.com), an online marketplace where respondents voluntarily enroll to complete questionnaires and other tasks. Using the MTurk platform, researchers have replicated laboratory results in, for example, economic games (including assessment of risk preferences, Eriksson and Simpson, 2010) (Fehr and Gächter, 2000; Mason and Suri, 2010) and social psychology tests (including framing effects, the conjunction fallacy and outcome biases) (e.g. Fagerlin et al., 2007; Paolacci et al., 2014; Rand, 2012). Mason and Suri (2012) provide a comprehensive discussion on the use of MTurk for behavioural research.

Respondents were told that the task would take approximately 20 minutes to complete and the payment was set at $2.50 (roughly consistent with the recommended reimbursement rate of USD$6/hour) (Figure 2-1). On MTurk all payments are processed through the Amazon.com intermediary which assures respondent anonymity is maintained. Respondents were asked to provide consent prior to beginning the study and to indicate that they understood that they were free to drop out of the study at any time without reason or penalty. An email address was given to respondents both at the beginning and the end of the questionnaire for any questions or concerns regarding their participation.



Health State Preference Survey- Instructions

Thank you for considering participating in this HIT.

We are conducting an academic survey about preferences for different health conditions.

You will be asked a series of questions about quality and length of life trade-offs and to provide reasons for your choices. You will also be asked to rate how you perceive your own health. If you are uncomfortable with doing so, please do not complete this HIT. The survey should take approximately 20 minutes to complete.

Figure 2-1: Screenshot of survey description on MTurk. HIT refers to a so-called 'Human Intelligence Task', in this instance, the Health State Preference Survey

2.42 Eligibility

Participation was limited to respondents located in Western Europe and North America in an attempt to limit potential cultural differences. There were no restrictions imposed in terms of respondent age, health status, or other demographic factors (e.g. education or employment status).

2.43 Questionnaire – Pilot

The questionnaire (described in detail in the next section) was piloted (n= 80) to test the feasibility of the MTurk platform. Particular items of interest were whether respondents were able to understand and complete the task in a time that reflected that adequate attention and effort was being dedicated, and to assess the feasibility of administering open-ended qualitative questions in support of TTO valuations.

Metrics such as the time each respondent spent per item and total time spent on the questionnaire suggested the sample was able to engage with the task and provide detailed responses to the questions they were asked. As an example of the care respondents appeared to invest in completing the questionnaire, one respondent noticed and reported an error in the display logic of the questions that had been overlooked during the questionnaire development.

Several modifications, namely the addition of a confirmation question to ensure respondents were providing their minimum number of years and minor rewording of questions, were undertaken based on pilot results. Further, based on the completion of open-ended response questions, it was decided that a structured ranking task (whereby possible response options are specified a priori) could be feasibly substituted with an open-ended format. An open-ended format is advantageous in that it does not confine respondents to certain options that may not truly reflect their decision and therefore potentially enhances the richness of the qualitative data gathered.

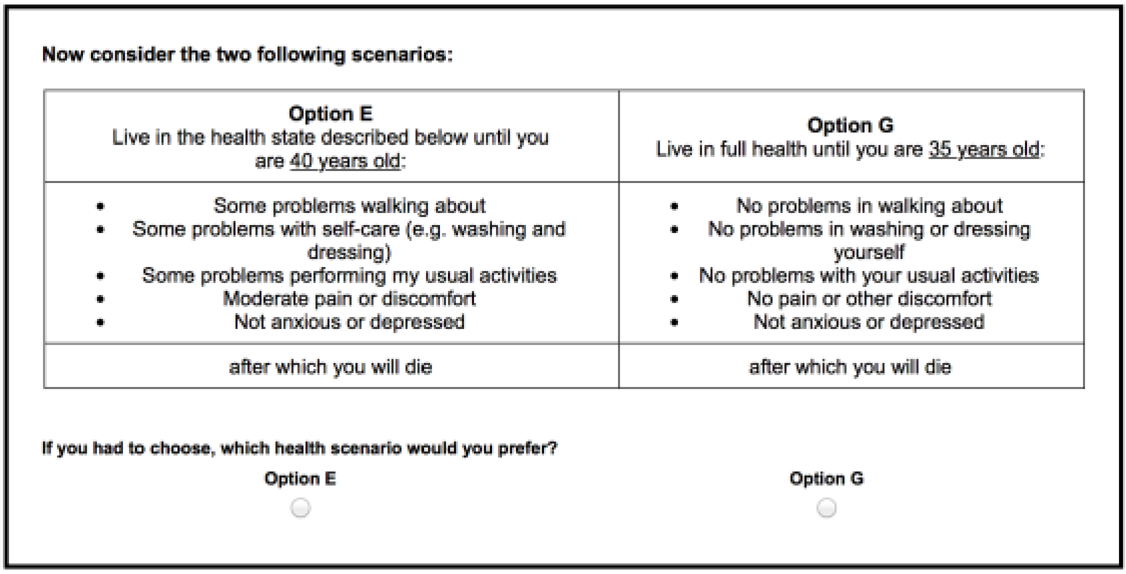
2.44 Questionnaire - Main Study

Demographic data were collected to characterize the sample. Information was gathered on gender, level of education, and experience with serious illness or injury (either the respondent themselves or someone close to them). Respondents were also asked to state their age prior to completing the TTO questions since this response was required for input into the age-framed questions for two of the groups. Respondents were asked to indicate their own perceived health using the EQ5D-5L health state descriptive system and a visual analogue scale (VAS). The EQ5D-5L is a generic preference-based measure of health. The EQ5D-5L describes HRQoL over five dimensions: mobility, self-care, usual activities, pain and anxiety using five different levels of functioning from “no problems,” “slight problems," "moderate problems," "severe problems" to "unable to" / "extreme problems". The VAS is a thermometer-type scale that requires respondents to indicate where health states lie on a scale ranging from 0 (worst health state imaginable) to 100 (best health state imaginable). Also, using the EQ5D framework, respondents indicated how they anticipated their health would be 30 years from now. At the end of the questionnaire, respondents were asked to state the age to which they expected to live (i.e. their subjective life expectancy or SLE) since asking them prior to the TTO questions may have primed them to use this age as an anchor in their valuations. For all of the questions in the survey, with the exception of the main TTO elicitations, respondents were given the option to decline to answer with a response option of "I prefer not to say."

Respondents were allocated to one of four groups, as seen in Figure 2-2. A two-by-two design was used involving two search procedures (matching and bisection) and two conditions (age-framed and standard). Two groups completed TTOs presented in a standard form. That is, for example, respondents were given the option between *x years in full health* and *10 years in the target health state*. The other two groups completed TTOs which were presented using an age-specific frame. Figure 2-3 shows an example of a 10-year age-framed TTO question. Note, a key benefit from using an online mode of administration for the age-frame is that age at death is computer-generated based on information on actual age provided by the respondent upon agreeing to participate.

TTO Frame	Search Procedure	Time Horizon (years) in TTO Questions
Standard	Matching	10, 30
Age-frame	Matching	10, 30
Standard	Bisection	10, 30
Age-frame	Bisection	10, 30

Figure 2-2: Main study design



Now consider the two following scenarios:

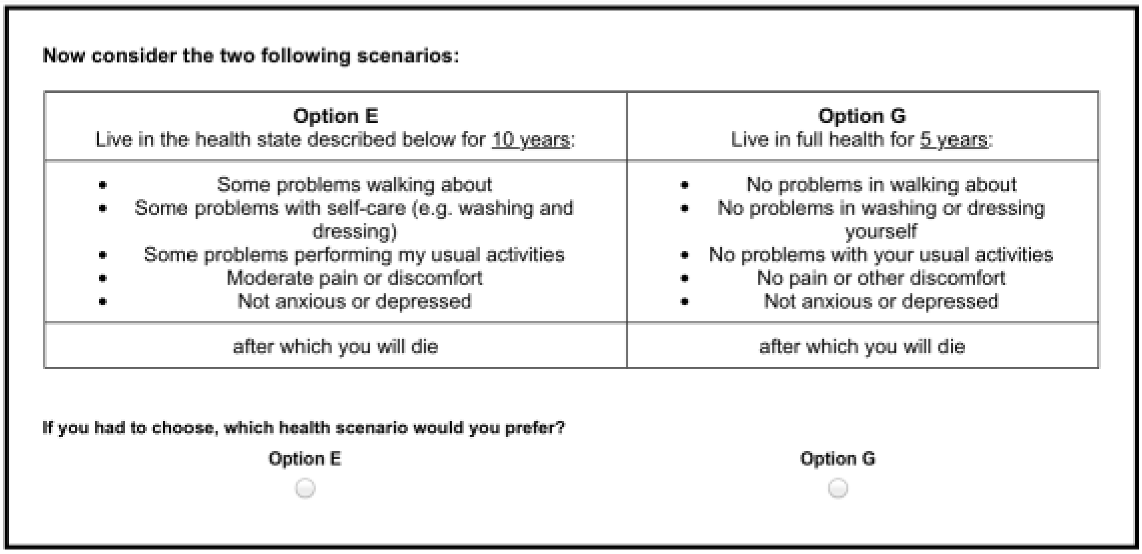
Option E Live in the health state described below until you are <u>40 years old</u>:	**Option G** Live in full health until you are <u>35 years old</u>:
• Some problems walking about • Some problems with self-care (e.g. washing and dressing) • Some problems performing my usual activities • Moderate pain or discomfort • Not anxious or depressed	• No problems in walking about • No problems in washing or dressing yourself • No problems with your usual activities • No pain or other discomfort • Not anxious or depressed
after which you will die	after which you will die

If you had to choose, which health scenario would you prefer?

Option E **Option G**

Figure 2-3: An age-framed 10-year TTO bisection question in which the respondent is 30 years old and they are asked whether they prefer five more years in full health to 10 years in the inferior health state.

Two groups used a matching search procedure and two groups used a choice-based (bisection) search procedure to complete the TTOs. When using the TTO to conduct health state valuations, there is no specific search procedure that is favoured. Choice-based techniques, commonly bisection, require the respondent to select among alternatives presented to them. The respondent is typically faced with a choice between x years in the health state (e.g. back pain) or some proportion (e.g. half of x years, hence bisection) in full health, as seen in Figure 2-4. Based on their response, the respondent makes a series of additional choices where the time horizon is further bisected until an indifference point is attained. By contrast, a matching search procedure, shown in Figure 2-5, requires the respondent to directly value one alternative in order to achieve indifference (i.e. equivalence) between the options without any prompting from the researcher. In this study for both matching and bisection, TTO valuations were elicited down to quarter-year increments. The iteration steps for the 10- and 30-year bisection questions can be found in Appendix 2-1.



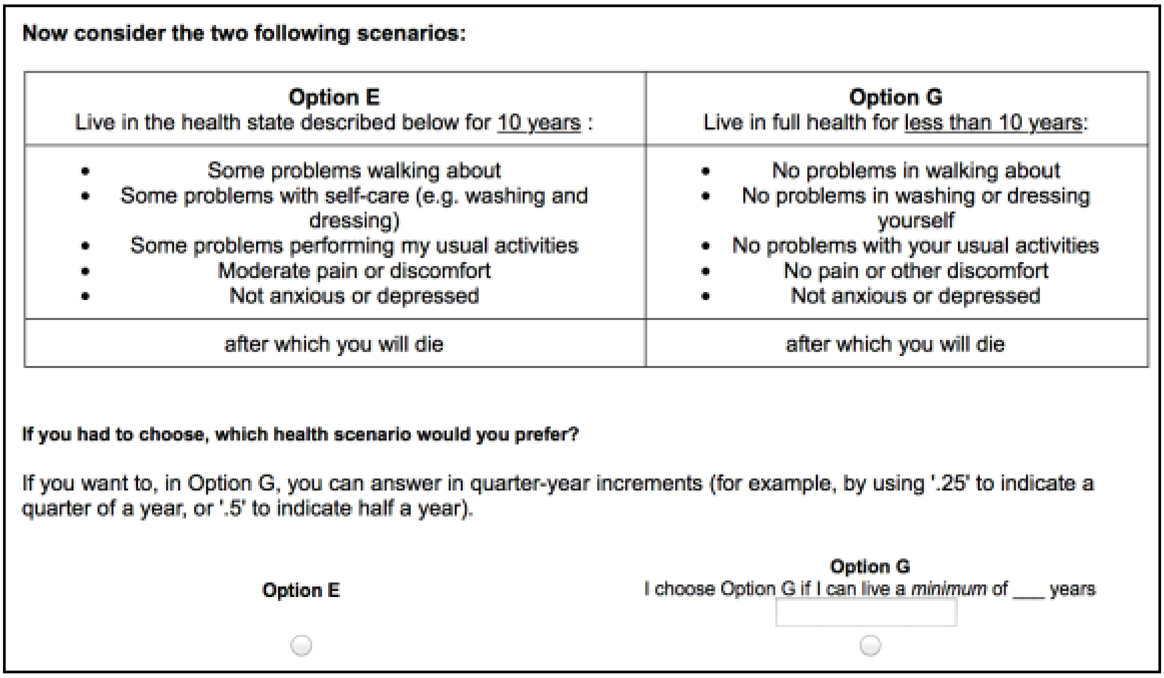
Now consider the two following scenarios:

Option E Live in the health state described below for 10 years:	**Option G** Live in full health for 5 years:
• Some problems walking about • Some problems with self-care (e.g. washing and dressing) • Some problems performing my usual activities • Moderate pain or discomfort • Not anxious or depressed	• No problems in walking about • No problems in washing or dressing yourself • No problems with your usual activities • No pain or other discomfort • Not anxious or depressed
after which you will die	after which you will die

If you had to choose, which health scenario would you prefer?

Option E **Option G**

Figure 2-4: Screenshot of an example of a standard 10-year bisection question



Now consider the two following scenarios:

Option E Live in the health state described below for 10 years :	**Option G** Live in full health for less than 10 years:
• Some problems walking about • Some problems with self-care (e.g. washing and dressing) • Some problems performing my usual activities • Moderate pain or discomfort • Not anxious or depressed	• No problems in walking about • No problems in washing or dressing yourself • No problems with your usual activities • No pain or other discomfort • Not anxious or depressed
after which you will die	after which you will die

If you had to choose, which health scenario would you prefer?

If you want to, in Option G, you can answer in quarter-year increments (for example, by using '.25' to indicate a quarter of a year, or '.5' to indicate half a year).

Option E

Option G
I choose Option G if I can live a *minimum* of ___ years

Figure 2-5: Screenshot of an example of a standard-frame 10-year matching question

Eliciting TTO values through matching and bisection enabled a test of procedural invariance to be carried out alongside the central test of descriptive invariance between the two frames. Procedural invariance, stipulates that preferences should be unaffected by the method through which they are elicited. Therefore, it should not matter whether a choice (e.g. bisection) or matching search procedure is used to elicit values. In the same way that a violation of descriptive invariance demonstrated by differences between the age-framed and standard conditions would give rise to questions of which values are more accurate, so would a violation of procedural invariance related to matching and choice search procedures.

There is reason to believe differences in values elicited using matching and bisection procedures would emerge, namely due to loss aversion and scale compatibility. Matching has been reported to be more susceptible to loss aversion, increasing values in matching tasks. For example, some authors have commented that matching highlights the fact that the respondent must *forgo* one attribute for the other (Attema and Brouwer, 2012). In contrast, choice-based procedures provide a more neutral decision context where the trade-off between attributes may be perceived to be of greater subtlety in a side-by-side comparison.

Another psychological phenomenon, scale compatibility, also affects attribute weighting. Scale compatibility refers to the tendency to assign greater weight to an attribute of an outcome that is consistent with the response scale in the task. In the TTO, respondents are asked to provide their answer in terms of duration; therefore, scale compatibility implies that more attention will be given to duration as opposed to health status. As a result, TTO values will be upwardly biased because the respondent will be less willing to trade-off life years for improvements in health status (Bleichrodt, 2002; Bleichrodt et al., 2003; Tversky et al., 1988). Similarly, Selart (1994) proposes that due to scale compatibility in matching tasks, reference (i.e. starting or anchor) points are implicitly provided whilst the same rationale cannot be applied to choice tasks. Note that one might also anticipate bisection would yield lower values than matching due to starting point bias since, in bisection, the iteration sequences are such that values start low then titrate upwards (e.g. Augestad et al. 2016; Ternent and Tsuchiya, 2013).

There were two main elicitation questions, a 10-year TTO and a 30-year TTO. The target health state (health state D, hereon termed HS_D and shown as Option E in figures 2-3, 2-4 and 2-5, above) was the same at both time horizons and was described using the EQ5D framework. The aim was to select a moderate health state such that respondents would be willing to make some trade-offs to avoid living in the state of poor health. If the health state was too serious, respondents may identify a period at which they find living in the state intolerable whereas if the health state was too mild, respondents may be unwilling to forgo any length of life to avoid it. In addition, van Nooten et al. (2014) used this health state, permitting comparisons to be made with their findings.

Prior to the main TTO questions, respondents were asked to rank HS_D along with full health (11111) and three other health states that were used in a practice TTO elicitation and in distraction questions. The purpose of the ranking exercise was to familiarize respondents with the health state descriptions and to have them contemplate how they might trade-off the different health state attributes (i.e. the severity of the different EQ5D dimensions) in a ‘timeless’ setting since, unlike TTO exercises, the health states were presented without a time horizon.

Following the ranking exercise, respondents were presented with a practice TTO elicitation to familiarize them to the format of the questions. In the matching condition, the practice question included a prompt to confirm that respondents had provided their minimum value in the TTO question. The prompt was only used in the matching condition (for both the standard and age-framed groups) and not in the bisection condition because of the nature of the bisection task which is assumed to guide the respondent to their lowest acceptable value through a series of iterative questions.

Respondent comprehension and attention was also assessed through a question involving choice dominance. In all four groups, in both the practice question and in the main TTO elicitations, the respondent was initially asked to indicate whether they preferred to live *x* years in HS_D or *x* years in full health. It is expected that the respondent will prefer the full health outcome for *x* years since it dominates the inferior health outcome for *x* years. If the respondent indicated that they preferred HS_D to full health for the same duration (*x* years) their data were excluded from the analyses.

Two time horizons were used in the main TTO elicitations, permitting tests of constant proportionality (CPTO), an underlying assumption of the TTO method which stipulates that a health state value should be the same regardless of the time horizon from which they are elicited. That is, it is assumed that the relative value of each individual year in a 40-year TTO, for example, will be the same as in a 10-year TTO so that if the respondent were to indicate indifference at 20 years in full health in the 40-year TTO for health state *x* (value = 20/40 or 0.50), they would (proportionally) opt for five years in the 10-year TTO for health state *x* (value = 5/10 or 0.50). Many studies have shown that respondents are generally more willing to trade duration for HRQoL at longer time horizons for several reasons (including, for example, decreasing marginal utility for later years; e.g. Essink-Bot et al., 2007; Martin et al., 2000; Stiggelbout et al., 1996). On the other hand, several studies have shown increasing TTO values at longer time horizons (Stalmeier et al., 1997; Unic et al., 1998). Given these results, the empirical validity of CPTO therefore remains in question.

2.45 Qualitative Data

In conjunction with the main TTO elicitations, qualitative data collection was undertaken with the aim of understanding how the age-frame impacts the types of considerations that respondents factor into their trade-off decisions compared to a standard presentation. Qualitative evidence enabled insight into the validity of the TTO values which is not available when looking at the values alone (e.g. whether the rationale underlying the decision is reflective of the health state or something else, such as the desire to reach a particular life goal regardless of health state severity).

Qualitative questions were open-ended and asked respondents to describe the types of things they were thinking about when they answered the preceding TTO question. 'Why' questions were avoided since some authors have commented that the underlying reasons for decisions are often difficult to express from this specific prompt given there are likely several complementary reasons contributing to a single decision (Patton, 2014). Therefore, in asking "what types of things" they considered, the respondent was free to provide as few or as many reasons as they felt were relevant. To complement the open-ended questions, respondents were asked to identify from a list of seven items including the five EQ5D dimensions (mobility, self-care, usual activities, pain or discomfort, anxiety or depression, length of time in HS_D, length of time in full health) the two aspects that had the most impact on their TTO response.

Throughout the questionnaire, the respondents completed a number of distractor questions which were aimed at spacing out the main TTO tasks and preventing them from recalling their answers from the previous trade-off questions. The distractor questions were simple tests of numeracy unrelated to health care decisions. These questions were selected on the basis that other studies have shown that respondents with low numeracy may be more susceptible to decision shortcuts and heuristics (and as a result, inconsistent preferences) (e.g. Reyna et al., 2009; Peters et al., 2006). These questions may shed light on this possible relationship. One such question asked the respondent to calculate how long it would take them to complete a trip based on a number of transit modes and layover times.

2.46 Feasibility Questions

After completing the main TTO elicitations, respondents were provided with a number of questions about their perception of the feasibility of the TTO task. Respondents were asked to rank a series of items relating to the amount and clarity of the information provided in the TTO tasks and more conceptual items relating to the sudden change to poor health, the ability to make trade-offs of short versus long durations, and the overall difficulty they had in completing the trade-offs.

2.47 Discounting

The TTO model assumes that respondents possess linear utility for duration (i.e. each incremental life year is equally valued) hence CPTO can be assumed. However, a substantial body of evidence suggests that respondents have positive time preference rates thereby assigning greater value to more proximal outcomes and discounting future outcomes. In order to take time preferences into account, individual discount rates were calculated based on the delay of ill health approach (Cairns, 1992). This approach requires respondents to imagine that at a certain time (Time A) they will enter a period in a given (inferior) health state lasting *x* days. The respondent is asked to provide the maximum number of days (y) in the given health state that they would be willing to accept if they could increase the period of delay from Time A to a given later time (Time B). Further details are provided in Appendix 2-2.

2.48 Analyses

Quantitative analyses were carried out in SPSS v21. Shapiro-Wilk Tests in both age- and standard frames showed that all eight distributions (two frames and two time horizons for matching and bisection groups respectively) deviated significantly from normality, and as a result nonparametric tests were used. The primary analysis of interest – whether TTO values in the age-framed condition were systematically different from TTO values elicited in the standard condition – was conducted using Mann-Whitney U tests. Mann-Whitney U tests assess the null hypothesis that independent groups (in this case, age-framed and standard sample) have the same median.

An a priori coding scheme was used to categorize qualitative data. In addition to the open-ended responses provided during the piloting of the questionnaire, the scheme was based on a review of the literature (e.g. Mulhern et al., 2013; Spencer, 2003) and, particularly, the findings of van Osch (2007) who used a think-aloud task to better understand how respondents arrived at their trade-off decisions. Relationships between TTO values and qualitative responses and whether any trends by search procedure or framing effects emerged were key points of interest.

The validity of the CPTO assumption was investigated using Wilcoxon Signed Ranks tests. Whether CPTO was observed more often depending on search procedure or frame was of interest. CPTO was also examined at an individual level since aggregate analyses may obscure important subgroup trends. Aggregate analysis of CPTO is the cumulative result of individual CPTO and may simply be an averaging out of individual response patterns that trend in opposing directions. Pliskin et al. (1980), for example, found that when examining trade-offs from 5- and 15-year time horizons, individual CPTO did not hold for most respondents; however, at an aggregate level, there was little difference in the mean values for each time horizon. A number of studies have shown that while CPTO is often violated at an individual level, aggregate findings offer more support for its validity (e.g. Sackett and Torrance (1978), Pliskin et al. (1980), Bleichrodt and Johannesson (1997), and Unic et al. (1998)). Some argue that the satisfaction of the CPTO requirement on an aggregate level is sufficient, this position assumes that aggregate preferences are ‘evening out’ variation between individuals, not that individuals have differently interpreted, or misinterpreted the task entirely (a question that can only really be assessed by looking at individual preferences).

Feasibility and importance rankings were analysed with Mann-Whitney U tests. The objective of the feasibility and importance questions was to gain insight in terms of whether presenting the TTO using an age-frame was more intuitive (i.e. easier to understand) as well as if it changed the weighting of attributes (i.e. length and HRQoL) when compared to responses from the standard TTO group (Mulhern et al., 2014).

Finally, Mann-Whitney U tests were used to assess whether response timing was related to TTO values. Response time may reflect the respondent’s engagement with the task or correlate with a particular pattern of TTO values. For example, respondents who adopt a decision heuristic (e.g. proportional responses to the time horizon, without consideration of health state severity) may complete the task faster than respondents who do not answer heuristically.

2.5 Results

Respondents were excluded post-hoc if they did not finish the questionnaire, completed the questionnaire in less than five minutes (this threshold was based on the distribution of completion times in the pilot questionnaire described below), had used a duplicate IP address[i], or did not check the consent box on the landing page. There were no substantive differences in the numbers of excluded respondents in the four groups. Table 2-1 contains the number of respondents excluded based on each criterion.

Table 2- 1: Respondents excluded from the analyses

	Matching		**Bisection**	
	Standard	Age-frame	Standard	Age-frame
Did not finish	8	5	9	8
Finished in less than five minutes	4	3	6	0
Duplicate IP address	0	3	1	0
Did not give explicit consent	2*	0	0	0

*Two respondents completed the questionnaire in its entirety; however, their data were excluded from the analysis.

In addition, when the TTO time horizons were equal, none of the respondents opted for the inferior health state over the full health outcome and therefore no respondents were excluded based on this dominance criterion.

2.51 The Sample

Questionnaires from a total of 432 respondents were analysed. The breakdown between the four respondent groups and their demographic characteristics are seen in Table 2-2. In terms of demographic characteristics, a single significant difference emerged between the four groups. The mean respondent VAS score (own health) was significantly different between the bisection standard and bisection age-framed groups (p=0.01).

Table 2-2: Respondent characteristics

Search Procedure:	**Matching**			**Bisection**		
	Standard Condition	Age-frame Condition	Standard vs Age-frame (p value)	Standard Condition	Age-frame Condition	Standard vs Age-frame (p value)
N	120	114		96	102	
Age, mean (range, SD)	38.5 (18-69, 12)	36.4 (20-69, 11)	0.08	37.4 (21-66, 12)	36.3 (18-65, 12)	0.40
Gender M : F (n)	51:69	50:64	0.83	40:56	50:52	0.30
EQ5D-5L own health, mean (range, SD)	0.86 (0.50-1, 0.11)	0.86 (0.18-1, 0.13)	0.37	0.83 (0.27-1, 0.15)	0.87 (0.39-1, 0.13)	0.10
VAS own health (0-100), mean (range, SD)	76 (25-100, 15)	75 (9-100, 17)	0.94	71 (19-100, 18)	77 (20-100, 15)	0.01
Experience with serious illness (personal or with someone close) (n)						
Yes	84	72		56	57	
No	35	40		37	41	
Prefer not to say	1	2		3	4	
Education (n)						
Some high school or less	0	0		2	2	
High school	42	44		41	34	
Undergraduate degree	68	57		46	54	
Advanced degree	10	10		7	12	
Prefer not to say	0	3		0	0	
SLE, mean (range, SD)	81 (50-110, 12)	78 (30-105, 13)	0.15	78 (35-125, 14)	80 (45-120, 13)	0.56
Remaining years of life, mean (range, SD)	42 (1-84, 16)	41 (4-76, 16)	0.88	41 (5-95, 17)	44 (6-96, 16)	0.35

Summary results for the four respondent groups are found in Table 2-3. Scatterplots of individual respondent values for VAS v. 10-year TTO and VAS v. 30-year TTO for both search procedures can be found in Appendix 2-4. Comparisons were made between the values in this study and both US and UK tariff values since the majority of the sample indicated they resided in the US. For matching, standard and age-framed mean values for HS_D were higher than the US tariff value of 0.738 and an even greater difference was seen higher than the UK tariff value of 0.648 (note this comparison contrasts matching used in this study and bisection used in the elicitation of tariff values). For bisection, the standard condition elicited the same TTO value as the UK tariff and the mean TTO value in the age-framed condition was substantially lower.

Table 2-3: VAS (Health State D) and 10- and 30-year TTO values

Search Procedure	Matching			Bisection		
Condition	Standard	Age-frame	*	Standard	Age-frame	*
N	120	114		96	102	
VAS for HS_D (0-100), mean (range, SD)	49 (18-85, 15)	45 (9-79, 14)	0.045	44 (5-92, 17)	48 (13-90, 15)	0.037
10-year TTO, median (IQR)	0.90 (0.80-1.00)	0.90 (0.82-1.00)	0.222	0.73 (0.48-0.88)	0.53 (0.28-0.83)	0.014
30-year TTO, median (IQR)	0.93 (0.83-0.99)	0.97 (0.84-0.99)	0.097	0.78 (0.38-0.92)	0.75 (0.35-0.92)	0.760

IQR = inter-quartile range.
* Standard vs. Age-frame, p-value

2.52 Within Matching Comparisons

VAS scores for HS_D, measured prior to TTO tasks (and therefore unrelated to the search procedure manipulation), were significantly different between the standard and age-framed conditions (p=0.045) (Table 2-3). No significant difference emerged between the standard and age-framed TTO values at either the 10-year or 30-year time horizon. Respondents were classified as non-traders if they were unwilling to forgo any length of life to avoid a lower HRQoL, therefore assigning the health state a value of one. Similar rates of non-trading were observed in both the age-framed and standard conditions in the 10-year TTO

(32 non-traders in the standard condition and 39 in the age-framed condition) and in the 30-year TTO (22 in the standard condition, 25 in the age-framed condition).

To shed light on possible response strategies, within-respondent trade-off patterns were investigated by looking at each individual's 10- and 30-year TTO values. Interestingly, approximately a quarter of respondents in each condition were willing to forgo the same number of years at both durations (23 respondents in the standard condition and 29 respondents in the age-framed condition). That is, if a respondent was willing to forgo two years in the 10-year TTO they were also willing to forgo two years in the 30-year TTO. This clearly has implications for individual level CPTO. Specifically, if the respondent was willing to forgo the same number of years in both the 10- and 30-year valuations, CPTO was not observed for these individuals (e.g. 8/10=0.80, 28/30=0.93). There was limited evidence of proportional trading in either frame for matching.

2.53 Within Bisection Comparisons

The mean VAS score for HS_D was significantly higher in the age-framed condition than in the standard condition (p=0.037). TTO values elicited in the standard condition were significantly higher than in the age-framed condition in the 10-year TTO (p=0.014) (Table 2-3). Standard and age-framed TTO values were similar at the 30-year time horizon. There were no non-traders (i.e. all respondents were willing to forgo *some* length of life for improved HRQoL) in both the standard or age-framed conditions in the 10-year TTO. In the 30-year TTO, there were nine and 13 non-traders in the standard and age-framed conditions, respectively.

Evidence of proportional trade-offs was present but infrequent: five respondents were willing to forgo roughly half the time horizon in both the 10- and 30-year TTOs, four respondents were willing to accept a third of the time horizon in either TTO and two respondents a quarter of either time horizon. Qualitative data offered additional insight. For instance, it was clear that some respondents in the standard bisection group relied on proportional heuristics ("…For this particular question I figured 3/4 of the possible time (7.5 vs. 10 years) in full health vs. some other lesser state of overall health was probably a fair trade", "I was considering whether it would be better to live half the time and feel well or to live twice as long but feel terrible. I chose health.").

2.54 Matching and Bisection Comparisons

A test of procedural invariance assessed differences between the search procedures within frames – i.e. between standard matching and bisection TTO values and between age-frame matching and bisection TTO values. Referring to Table 2-3, 10- and 30-year standard TTO values differed significantly between matching and bisection ($p<0.001$), as did age-framed standard and bisection TTO values ($p<0.001$).

2.55 Constant Proportional Trade-Off

CPTO is a necessary assumption of the TTO model which, put simply, ensures that health state values can be considered independent from the horizon from which they are elicited. On an aggregate level, Wilcoxon-Rank tests showed proportionality in all four respondent groups, i.e. the numbers of respondents in each group for whom the difference between 10- and 30-year TTO values was positive was similar to the number for whom it was negative.

As discussed above, there was evidence of non-compensatory response strategies on the part of several respondents (e.g. where the respondent has traded-off the same number of years at both time horizons). Table 2-4 shows the number of respondents and percentage in each group for whom the difference between their 10-year and 30-year TTO values was between -0.10 and 0.10 (a relaxed definition of CPTO[ii]), and the distribution of respondents with greater variability. The majority of respondents in each group (except the age-frame bisection group) provided values consistent with CPTO according to this relaxed definition, although proportionality did not hold for at least a quarter of individual responses in each group. Importantly, qualitative evidence was able to shed light on the reasons why some respondents' values might feign CPTO (e.g. proportional heuristics) or to clearly indicate different valuation strategies (e.g. basing trade-offs on different life goals or forgoing the same number of years in both the 10- and 30- year TTOs). This comparison between individual and aggregate TTO findings highlights the importance of both types of analyses as well as qualitative data in better understanding valuation strategies and whether the assumption of CPTO is met.

Table 2-4: Distribution of individual differences between 30-year TTO and 10-year TTO values

		-1.00 to -0.91	-0.90 to -0.81	-0.80 to -0.71	-0.70 to -0.61	-0.60 to -0.51	-0.50 to -0.41	-0.40 to -.0.31	-0.30 to -0.21	-0.20 to -0.11	-0.10 to 0.10 (%)	0.11 to 0.20	0.21 to 0.30	0.31 to 0.40	0.41 to 0.50	0.51 to 0.60	0.61 to 0.70	0.71 to 0.80	0.81 to 0.90	0.91 to 1.00
Matching	Standard	0	0	0	0	0	1	3	2	14	76 (72%)	7	0	0	0	1	0	0	0	1
	Age-Frame	0	0	0	0	1	0	0	2	6	76 (78%)	8	3	1	1	0	0	0	0	0
Bisection	Standard	0	0	0	0	1	2	3	1	11	59 (63%)	10	3	2	1	0	0	0	0	0
	Age-frame	1	0	0	0	0	0	4	4	4	48 (48%)	13	10	3	8	3	2	0	0	0

*n values are less than full samples due to a small number of respondents providing responses that were not logical (e.g. when the question asked for fewer years than the time horizon, they provided a greater number of years from the time horizon) as well as some non-responses for either 10- or 30- year TTOs.

2.56 Importance Rankings

Table 2-5 presents the results of the importance ranking exercise where respondents were asked to select the two aspects of the health states that most affected their valuation. In brief, the purpose of this exercise was to better understand if there was a particular aspect of the TTO to which the respondent was allocating a significant amount of weight in their decision. When respondents used a matching procedure, the mobility dimension was most often ranked as the most important aspect in both standard (29/120) and age-framed (33/114) conditions. On the other hand, when values were elicited through bisection, the most commonly cited aspects were 'time spent in HS_D' in the standard condition (26/96) and 'pain and discomfort' in the age-framed condition (22/102).

Table 2-5: Ranking of importance of aspects of the health state in TTO valuation

	Matching				Bisection			
Condition	Standard N=120		Age-frame N=114		Standard N=96		Age-frame N=102	
Response ranking	***1st***	***2nd***	***1st***	***2nd***	***1st***	***2nd***	***1st***	***2nd***
Mobility	29	17	33	15	10	8	17	10
Self-care	27	10	16	21	14	9	17	18
Usual activities	6	15	3	10	0	7	4	16
Pain, discomfort	26	25	19	25	15	20	22	13
Anxiety, depression	11	20	13	17	9	11	11	8
Time in HS_D	8	17	10	11	26	14	17	22
Time in full health	21	18	26	18	22	27	17	21

2.57 Qualitative Results

The qualitative data gathered after each main TTO elicitation question were coded into seven categories (Table 2-6). Respondents were asked to describe ‘the kinds of things they were thinking about’ when completing the preceding TTO. In some instances, more than a single theme was found in a response; however, only one theme was recorded for each respondent. Descriptions of the themes can be found in Appendix 2-3.

Table 2-6: Qualitative responses

10-year TTO	Matching		Bisection	
	Standard	Age-frame	Standard	Age-frame
HRQoL vs Longevity	36	32	31	32
Strong preference for length of life, longevity-related goals (e.g. family)	33	32	39	41
HS effects on self (health goals, e.g. activities, enjoyment)	27	29	7	15
HS effects on others (e.g. burden)	7	5	10	3
MET	1	1	2	3
Unclear or none	10	15	5	15
30-year TTO	Matching		Bisection	
	Standard	Age-frame	Standard	Age-frame
HRQoL vs Longevity	49	38	36	34
Strong preference for length of life, longevity-related goals (e.g. family)	22	34	37	29
Health effects on self (health-goals, e.g. activities, enjoyment)	20	15	10	12
HS effects on others (e.g. burden)	7	4	2	6
MET	7	1	7	2
Unclear or none	15	19	4	26

Based on their qualitative responses, only about a third of respondents in each of the four groups seem to have completed their TTO valuations by counterbalancing length of life and HRQoL. Several other significant decision rationales emerged amongst similarly large proportions of respondents, raising questions as to the extent to which values capture the severity (i.e. the HRQoL) of HS_D in a valid way. The qualitative data revealed that the perceived severity of HS_D varied widely between respondents with some commenting that the health state would be manageable:

> We all have days or moments in life when we do not feel well but that does not mean we would rather be dead- so, if I had to feel not so great every day, I would still want to live as long as I could get around some and have help with the things I needed help with. (Standard matching TTO respondent)

Others felt that the health state would be almost intolerable ("The longer you live a suffering life, the unhappier you will be so sometimes it is better to go earlier, without the suffering" (age-frame matching TTO respondent)). To this end, several respondents indicated that they did not wish to die while in poor health and there is some evidence in the literature that suggests that respondents have preferences for particular patterns in their health over time guided by gestalt-type characteristics (e.g. trends, peaks, and ends) (Oliver, 2008). In addition, the observation of pain as an important factor in TTO decisions, also found in the importance rankings, was corroborated through the qualitative data, for example "I just rather live without pain no matter how long that is" (standard matching TTO respondent), and:

> I was thinking how much life of comfort and good health is worth. For me, living a life in pain and without the ability to enjoy life, why would it be worth living more years. I would rather live fewer years with the capacity to fully enjoy myself that to live a reduced life. (Standard matching TTO respondent)

This finding is consistent with several other studies which have reported a large impact of pain/discomfort on valuation (Borgström et al., 2012; Brazier et al., 2011).

2.58 Feasibility questions

The results of the feasibility questions are found in Table 2-7. Comparing search procedures and frames on perceived feasibility was done with the objective of identifying reasons for preferring one TTO presentation over another. When making comparisons according to search procedure (grouping standard and age-framed conditions together), notable differences emerged between bisection and matching on two items.

It was difficult to draw firm conclusions in regards to group differences in terms of feasibility. Comparisons between the standard and age-framed conditions (i.e. standard matching vs. age-framed matching and standard bisection versus age-framed bisection) revealed that the four groups were remarkably similar on all items. Considering three items in particular – the amount of information included in the HS description, chance of relief and the description of full health – the majority of respondents in all four groups appear to have understood the main parameters of the TTO task.

Table 2-7: Perceived feasibility* of the TTO

Search Procedure	Matching							Bisection						
Condition	Standard			Age-frame			P-value, Standard vs. age-frame	Standard			Age-frame			P-value, Standard vs. age-frame
	Agree	Not really	Disagree	Agree	Not really	Disagree		Agree	Not really	Disagree	Agree	Not really	Disagree	
Too much information in HS descriptions	12	43	65	16	33	65	0.915	13	31	52	5	43	53	0.736
Full health not clear	11	17	92	10	18	86	0.856	7	14	75	8	15	78	0.872
Difficult to imagine the change	28	42	50	30	34	50	0.993	21	33	42	29	40	32	0.086
Chance of relief	18	23	79	20	24	70	0.473	11	12	73	13	30	58	0.014
Decision easier with long time horizon	59	39	22	51	44	19	0.691	57	25	14	57	26	18	0.601
Considered how others might decide	26	28	66	14	36	64	0.451	18	21	57	14	32	55	0.784
Shorter time in full health not realistic	43	34	43	44	28	42	0.874	28	34	34	32	34	35	0.781
Thought how HS would affect responsibilities	82	22	16	81	16	17	0.762	65	15	16	65	24	12	0.857
In general, difficult to answer	31	49	40	23	47	44	0.278	23	33	40	28	37	36	0.381

*Adapted from Mulhern et al. (2014)

Approximately half of the respondents in each group indicated that they found it easier to make a trade-off at the longer, 30-year time horizon. Comparable response patterns were observed across all four groups in terms of question difficulty, descriptions of full health and HS_D and difficulty in perceiving the sudden change to poor health.

2.59 Questionnaire completion times

On an aggregate level, mean response times were compared between the four respondent groups. Other studies have observed that the time it takes respondents to complete the survey or questionnaire is dependent on the search procedure (e.g. Lenert et al., 1998). Also, response times may provide some insight into engagement levels. For example, if a particular presentation of the task is more difficult to understand, respondents in this condition may take longer to complete the questions since they need to think carefully about the information presented. As seen in Table 2-8, no significant differences emerged in any comparisons.

Table 2-8: Questionnaire completion times

	Standard mean (in minutes) (range, SD)	Age-Frame mean (in minutes) (range, SD)	
Matching	15.13 (6-41, 6.02)	16.23 (5-38, 6.55)	p=0.192
Bisection	15.42 (6-40, 6.03)	17.03 (5-59, 9.16)	p=0.898
	p=0.742	p=0.545	

2.6 Discussion

The initial hypothesis that an age-framed TTO would yield higher values than a standard TTO received little support in this study. Instead, two unexpected results were observed which are deserving of further investigation. First, the age-frame had no significant effect on values when a matching procedure was used. Second, when a bisection procedure was used, *lower* values were elicited in the age-frame condition than in the standard condition.

Beginning with the matching results, no significant difference in TTO values emerged between the standard and age-framed conditions. The matching results are therefore inconsistent with the findings of van Nooten et al. (2014) who found that when respondents were reminded about their age at death in a 10-year TTO they were less willing to forgo longevity for improved HRQoL. One-quarter of respondents in age and standard groups were willing to forgo the same number of years in both the 10- and 30-year TTOs providing some evidence suggesting this segment of respondents may have relied on rules of thumb to reach their valuation. That respondents were using a decision strategy of this nature is a possible reason why CPTO doesn't consistently hold.

Interestingly, when a bisection search procedure was used, the age-frame had a significant impact in the opposite direction than predicted: age-framed values were lower than standard values. Although it was anticipated that using the respondent's age at death as the time horizon would elicit a stronger effect of loss aversion, it seems that the frame may have increased the amount of attention directed towards the HRQoL attribute. It may be the case that respondents have a particularly strong focus on the length of life attribute in the standard condition since, in order to compute their age at death, their attention must be directed towards this attribute (i.e. the time horizon). Two examples of qualitative comments are: “I was wondering how old I'll be when I die. I was also remembering what it was like to be able to walk around with no problems” (standard matching respondent), and “I was thinking about how old I'll be in 20-30 years, what I plan to be doing and how age will affect and be affected by my general health. That is, certain things will be trickier” (standard bisection respondent). Adopting this interpretation, in comparison to the standard TTO, the age-framed condition frees up some of the respondent’s cognitive budget, allowing the respondent to focus relatively greater attention towards the health state. Matza et al. (2015) suggests that in the context of the LE TTO respondents are better able to focus on the health state because they’re not distracted by an unrealistic time horizon. It may also be that respondents in standard choice-based tasks are using similar decision frameworks as those used by respondents in the age-frame condition.

The directional influence of the age-frame TTO was not as hypothesized, and was inconsistent over the 10- and 30-year time horizons. However, the greatest difference between standard and age-frame occurs at the 10-year time horizon. This time horizon is frequently adopted and, importantly, is that recommended by the NICE reference case. Therefore, it would be useful to conduct further research which looks at how question framing, in terms of age versus standard TTOs, impacts thought processes and, ultimately, TTO valuations (including what the relationship between age-frame and standard TTO values looks like over different time horizons). In practical terms, if the age-frame does indeed enable respondents to more easily internalize the TTO then this method might be considered for more widespread implementation.

2.61 Qualitative findings

The majority of qualitative responses were quite detailed, lending support to online self-complete valuation tasks as a viable method to elicit preferences for health states. It is interesting to note that the qualitative coding yielded similar patterns across all four groups. It seems that, although trade-off decisions are based on similar reasoning across groups, behavioural economic influences of which the respondent is unaware (e.g. scale compatibility, loss aversion) have varying effects on TTO values depending on the frame and search procedure.

Many respondents cited reasons underlying their trade-off decisions that were unrelated to the severity of the health state. Typically, the effects of a health intervention on anyone other than the individual are excluded from economic evaluation. However, an increasing amount of research is concerned with measuring the social costs of caring (e.g. physical/emotional strain, wellbeing and/or opportunities foregone by others) (e.g. Arnesen and Trommald, 2005; Basu and Meltzer, 2005; van Nooten et al., 2016). For instance, a number of respondents expressed a strong willingness to live a particular length of time for reasons relating to spending more time with family and loved ones and caring for others. Bobinac (2012) referred to a caregiving effect and a family effect, for which there was evidence in this study ("I want to live long enough to see my kids become grown and independent people… to see them get married and help with their children" (age-framed matching respondent). Whether or not the health state compromised their ability to care for others was also mentioned by a number of respondents.

If extrinsic goals form the basis of respondent decisions under certain circumstances (e.g. when the time horizon is short), future research might investigate multiple health states since it could be expected that different health states are assigned the same TTO value based on the common goal (Van der Pol and Shiell, 2007).

2.62 Feasibility

Deciding on the appropriate time horizon to use in the TTO is a longstanding and ongoing point of discussion. The commonly implemented 10-year time horizon has been criticized on the grounds of being rather artificial since many respondents anticipate living much longer than 10 years. To this extent, the feasibility questions revealed that approximately half of the respondents in each of the four groups thought that the trade-off at the longer duration was more easily undertaken. This finding may be due in part to the subjective life expectancies (and thus remaining years) of respondents in this study, in that the time horizon in the 30-year TTO more closely coincided with their external expectations for their age at death. Although longer durations may be perceived as more feasible, by increasing the time horizon, potential issues such as maximum endurable time (i.e. where the individual reaches a point in which they no longer positively value longevity in the given health state), discounted utility (whereby the respondent values each subsequent year of life less and less), and lower anticipated HRQoL at older ages, all of which would have a deflating effect on TTO values, must be addressed.

2.7 Limitations

Several limitations to the study should be noted. First, the questions were asked in the same order throughout, specifically the 10-year TTOs were completed before the 30-year TTOs. In addition, the main valuation task focused on a single health state (HS_D). Other studies have shown that the valuation of a health state can be affected by the other health states valued at the same time and that the order of the health states influences the values that are elicited (Pinto-Prades, 2013). Future research may wish to examine the effect of the age-frame when valuing multiple health states and randomizing the presentation of the TTO exercises between respondents.

Finally, the study was carried out through online self-complete questionnaire. While self-complete online questionnaires are becoming increasingly used throughout the literature and provide a number of benefits (e.g. in terms of paradata such as time spent per question or number of clicks per screen, and, when compared to typical student populations, a wider spectrum of illness/experience/age), there are also notable shortcomings to this approach. For instance, Craig et al. (2014) note that web-based surveys may exclude particular subgroups of the population, such as underserved groups without access to a computer or the Internet. As is almost inevitable with experiments of this nature (whether it is in academic labs or online), some level of self-selection is unavoidable. Given the importance to any result of establishing external validity, future research should seek to validate the results found in this study by, for example, using face-to-face interviews.

2.8 Broader Implications

The present study contributes to the literature on both practical and conceptual levels in drawing attention to the importance of *how* we elicit values for implementation into economic analyses. In practical terms, the study adds to a growing body of research which uses self-complete web-based questionnaires for health state valuation exercises. Online platforms offer several interesting differences when compared to traditional studies evaluating methodological aspects of health state valuation elicitation tools, which often rely on student samples. Specifically, web-based studies may also allow for a wider respondent pool than those often found in convenience samples or samples of student respondents.

Drawing upon a larger and more diverse range of observations than previously reported in the literature, this study found a wide spectrum of health status among the respondents and also rich qualitative evidence which allowed a number of potential influences on TTO values to be identified. Although there are advantages and disadvantages to any method of administration, this study showed that a self-complete format may provide a useful means through which hypotheses can be tested. Crucially, the study raises a number of questions in regards to the validity of a number of the fundamental assumptions underlying the TTO model, exemplified through violations of procedural invariance between search procedures and across frames for both 10- and 30-year TTOs (see Table 2-9).

Table 2-9: Review of hypotheses and results

Test	Hypothesis	Result
Descriptive invariance	*An age-framed TTO yields higher values than a standard TTO*	The age-frame had no significant effect on values when a matching procedure was used. However, when a bisection procedure was used, *lower* values were elicited in the age-frame condition than in the standard condition.
Procedural invariance	*Matching values are expected to be higher than bisection values for both standard and age-framed TTOs due to effects of loss aversion and scale compatibility*	10- and 30-year standard TTO values differed significantly between matching and bisection ($p<0.001$), as did age-framed standard and bisection TTO values ($p<0.001$).
CPTO	*TTO values should not differ significantly depending on the time horizon from which they are elicited*	CPTO held across all comparisons

From a conceptual perspective, the study sought to examine the effect of framing the time horizon in a logically equivalent way as is standardly done, implementing eventual age of death as the time horizon. It was thought that given the abstract nature of the time horizon presented in a standard TTO, respondents may fail to appreciate the limitation placed on their life expectancy. Essentially, it draws attention to the question: if respondents aren’t compensating gains in HRQoL with losses in length of life, then what are they considering?

Heterogeneity in how respondents interpret the time horizon invalidates the comparability of values and their aggregation in welfare estimates and may also help to explain violations of TTO assumptions found in the literature (Collins, 2003). For instance, van Nooten et al. (2015) found that some respondents expressed lack of awareness that a 10-year TTO time horizon often drastically reduced their remaining life span, lending support to the possibility that these respondents use the time frame in some other (unknown) manner. Responses in the present study, in particular the qualitative findings, provide evidence that varying interpretations of the time horizon are a source of inter-respondent variability. For instance, some respondents clearly answered by considering objectives unrelated to the health state whereas others used more crude, proportional strategies, apparently deriving little meaning from either the time horizon or the health state.

It is important to note that small differences in health state valuations have the ability to modify the outcome of cost-utility models (Naglie and Detsky, 1992; Pignone et al., 2007). For example, as was shown in this study, even if the difference is seemingly small, this could have large implications in terms of cost-effectiveness analyses. Small differences in values assigned to a health state are particularly amplified when a large number of patients is considered over a long time-span as, for example, in chronic disease (Matza et al., 2015).

Notably, Oliver and Wolff (2014) explain that if all values are biased in the same direction, such biases are relatively unproblematic since the cost-effectiveness threshold itself will then be set at an artificially high level. It is important to consider, however, that potentially invalidating issues relating to comprehension (e.g. heterogeneity in understanding or interpretation) should be considered and that these issues would argue in favour of a more age based/intuitively-framed task. That is, if the framing of the question influences response strategy, the primary concern becomes less about values being relatively higher and lower but rather the reasoning underlying the valuation: is one method for eliciting values more valid and gives a better reflection of the health state – than another method?

To establish consistency with the methods laid out by organizations such as the EuroQoL Group, researchers most often implement a 10-year time horizon. A number of concerns have been raised in the literature in terms of the feasibility of this short expectation for longevity – often perceived as artificially short. This study lends support to specific concerns already raised in terms of comprehension of the TTO and violations of procedural invariance, among other issues, and it draws attention to important additional issues which require further investigation and understanding if the health state values generated using the TTO are to be considered valid. The trade-off between longevity and HRQoL presented to respondents in the TTO is a difficult decision to make so researchers should strive to make it as intuitive as possible for respondents. Doing so is in the best interests of both individuals who are presented with this type of decision - whether it is for elicitation purposes or in real world clinical treatment choices and of policy-makers who seek to implement valid health state quantifications into economic analyses.

References

Amir, O., & Rand, D. G. (2012). Economic games on the internet: The effect of $1 stakes. *PloS One, 7*(2), e31461.

Arnesen, T., & Trommald, M. (2005). Are QALYs based on time trade-off comparable? - A systematic review of TTO methodologies. *Health Economics, 14*, 39-53.

Attema, A. E., & Brouwer, W. B. F. (2012). The way that you do it? An elaborate test of procedural invariance of TTO, using a choice-based design. *The European Journal of Health Economics, 13*(4), 491-500. doi: 10.1007/s10198-011-0318-y

Augestad, L. A., Rand-Hendriksen, K., Stavem, K., & Kristiansen, I. S. (2013). Time trade-off and attitudes toward euthanasia: implications of using 'death' as an anchor in health state valuation. *Quality of Life Research, 22*. doi: 10.1007/s11136-012-0192-9

Augestad LA, Stavem K, Kristiansen IS, Samuelsen CH, Rand-Hendriksen K. (2016) Influenced from the start: anchoring bias in time trade-off valuations. *Quality of Life Research*. 25:2179-2191. doi:10.1007/s11136-016-1266-x.

Basu, A., & Meltzer, D. (2005). Implications of spillover effects within the family for medical cost-effectiveness analysis. *J Health Econ*, *24*(4), 751-773. doi: 10.1016/j.jhealeco.2004.12.002

Bleichrodt, H. (2002). A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Economics, 11*, 447-456.

Bleichrodt, H., Luis Pinto, J., & Abellan-Perpinan, J. M. (2003). A consistency test of the time trade-off. *Journal of Health Economics, 22*, 1037-1052.

Bleichrodt, H., & Johannesson, M. (1997). Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. *Journal of Health Economics, 16*(2), 155-175.

Bobinac, A. (2012). *Economic evaluations of health technologies: insights into the measurement and valuation of benefits.* Erasmus University Rotterdam, The Netherlands.

Borgström, F., Aghayev, E., Olafsson, G., & Miltenburger, C. PMD77 Contribution of Individual EQ-5D Dimensions to Improved Quality of Life After Balloon Kyphoplasty for Vertebral Compression Fractures. *Value in Health, 15*(7), A358-A359. doi: 10.1016/j.jval.2012.08.926

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health, 27*(3), 281-291. doi: 10.1093/pubmed/fdi031

Brazier, J., Rowen, D., Tsuchiya, A., Yang, Y., & Young, T. A. (2011). The impact of adding an extra dimension to a preference-based measure. *Soc Sci Med, 73*(2), 245-253. doi: 10.1016/j.socscimed.2011.05.026

Brouwer, W., & van Exel, N. J. A. (2005). Expectations regarding length and health related quality of life: Some empirical findings. *Social Science and Medicine, 61*, 1083-1094.

Buitinga, L., Braakman-Jansen, L. M. A., Taal, E., & van de Laar, M. (2012). Construct validity of the interview Time Trade-Off and computer Time Trade-Off in patients with rheumatoid arthritis: A cross-sectional observational pilot study. *BMC Musculoskelet Disord, 13*, 112. doi: 10.1186/1471-2474-13-112

Cairns, J. A. (1992). Health, wealth and time preference. *Project Appraisal, 7*(1), 31-40. doi: 10.1080/02688867.1992.9726836

Chapman, G. B., Elstein, A. S., Kuzel, T. M., Nadler, R. B., Sharifi, R., & Bennett, C. L. (1999). A multi-attribute model of prostate cancer patients' preferences for health states. *Quality of Life Research, 8*(3), 171-180.

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research, 12*(3), 229-238. doi: 10.1023/a:1023254226592

Craig, B. M., & Busschbach, J. J. (2009). The episodic random utility model unifies time trade-off and discrete choice approaches in health state valuation. *Population Health Metrics, 7*(1), 1-10. doi: 10.1186/1478-7954-7-3

Craig, B. M., Reeve, B. B., Brown, P. M., Cella, D., Hays, R. D., Lipscomb, J., . . . Revicki, D. A. (2014). US Valuation of Health Outcomes Measured Using the PROMIS-29. *Value in Health, 17*(8), 846-853. doi: http://dx.doi.org/10.1016/j.jval.2014.09.005

Dolan, P., & Stalmeier, P. (2003). The validity of time trade-off values in calculating QALYs: constant proportional time trade-off versus the proportional heuristic. *Journal of Health Economics, 22*(3), 445-458.

Essink-Bot, M.-L., Stuifbergen, M., Meerding, W.-J., Looman, C., Bonsel, G., & Group, V. (2007). Individual differences in the use of the response scale determine valuations of hypothetical health states: an empirical study. *BioMed Central Health Services Research, 7*(62).

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D.

M. (2007). Measuring Numeracy without a Math Test: Development of the Subjective Numeracy Scale. *Medical Decision Making, 27*(5), 672-680. doi: 10.1177/0272989x07304449

Fehr, E., & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *The Journal of Economic Perspectives, 14*(3), 159-181.

Franke, G. R. (2010). Multicollinearity *Wiley International Encyclopedia of Marketing*: John Wiley & Sons, Ltd.

Gigerenzer, G. (2002). *Adaptive Thinking: Rationality in the Real World*: Oxford University Press.

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics, 14*(3), 399-425.

Hutchins, R., Viera, A. J., Sheridan, S. L., & Pignone, M. P. (2015). Quantifying the Utility of Taking Pills for Cardiovascular Prevention. *Circulation: Cardiovascular Quality and Outcomes*. doi: 10.1161/circoutcomes.114.001240

Lenert, L. A., Cher, D. J., Goldstein, M. K., Bergen, M. R., & Garber, A. (1998). The effect of search procedures on utility elicitations. *Medical Decision Making, 18*(1), 76-83.

Martin, A., Glasziou, P., Simes, R., & Lumley, T. (2000). A comparison of standard gamble, time trade-off, and adjusted time trade-off scores *International Journal of Technology Assessment in Health Care, 16*(1), 137-147.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods, 44*(1), 1-23.

Matza, L. S., Boye, K. S., Feeny, D. H., Bowman, L., Johnston, J. A., Stewart, K. D., . . . Jordan, J. (2015). The time horizon matters: results of an exploratory study varying the timeframe in time trade-off and standard gamble utility elicitation. *Eur J Health Econ*. doi: 10.1007/s10198-015-0740-7

McFarlane, P. A., Pierratos, A., Bayoumi, A. M., & Redelmeier, D. A. (2007). Estimating Preference Scores in Conventional and Home Nocturnal Hemodialysis Patients. *Clinical Journal of the American Society of Nephrology, 2*(3), 477-483. doi: 10.2215/cjn.03941106

Mulhern, B., Bansback, N., Brazier, J., Buckingham, K., Cairns, J., Devlin, N., ... & Rowen, D. (2014). Preparatory study for the revaluation of the EQ-5D tariff: methodology report. *Health Technology Assessment,* 18(12).

Mulhern, B., Longworth, L., Brazier, J., Rowen, D., Bansback, N., Devlin, N., & Tsuchiya,

A. (2013). Binary choice health state valuation and mode of administration: head-to-head comparison of online and CAPI. *Value in Health, 16*(1), 104-113.

Naglie, I. G., & Detsky, A. S. (1992). Treatment of chronic nonvalvular atrial fibrillation in the elderly: a decision analysis. *Med Decis Making, 12*(4), 239-249.

Norman, R., King, M. T., Clarke, D., Viney, R., Cronin, P., & Street, D. (2010). Does mode of administration matter? Comparison of online and face-to-face administration of a time trade-off task. *Quality of Life Research, 19*(4), 499-508.

Oliver, A. (2008). Assessing the influence of gestalt-type characteristics on preferences over lifetime health profiles. *Medical Decision Making, 28*(5), 723-731.

Oliver, A., & Wolff, J. (2014). Are people consistent when trading time for health? *Econ Hum Biol, 15*, 41-46. doi: 10.1016/j.ehb.2014.05.001

Paolacci, G., & Chandler, J. (2014). Inside the turk. Understanding mechanical turk as a participant pool. *Current Directions in Psychological Science, 23*(3), 184-188.

Patton, M. Q. (2014). *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*: SAGE Publications.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science, 17*(5), 407-413.

Pignone, M., Earnshaw, S., Pletcher, M. J., & Tice, J. A. (2007). Aspirin for the primary prevention of cardiovascular disease in women: a cost-utility analysis. *Arch Intern Med, 167*(3), 290-295. doi: 10.1001/archinte.167.3.290

Pinto-Prades, J. L. (2013). Imprecise preferences and order effects in the time trade-off.

Pliskin, J. S., Shepard, D. S., & Weinstein, M. C. (1980). Utility Functions for Life Years and Health Status. *Operations Research, 28*(1), 206-224. doi: 10.1287/opre.28.1.206

Rand, D. G. (2012). The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology, 299*, 172-179.

Rencz, F., Brodszky, V., Péntek, M., Bereczki, D., & Gulácsi, L. (2015). Health state utilities for migraine based on attack frequency: a time trade-off study. *Neurological Sciences, 36*(2), 197-202.

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How Numeracy Influences Risk Comprehension and Medical Decision Making. *Psychological bulletin, 135*(6), 943-973. doi: 10.1037/a0017327

Richardson, J., Iezzi, A., Khan, M. A., & Maxwell, A. (2014). Validity and Reliability of the Assessment of Quality of Life (AQoL)-8D Multi-Attribute Utility Instrument.

The Patient, 7(1), 85-96. doi: 10.1007/s40271-013-0036-x

Sackett, D., & Torrance, G. (1978). The utility of different health states as perceived by the general public. *Journal of Chronic Diseases, 31*(11), 697-704.

Selart, M. (1994). Preference reversals in judgment and choice: the prominence effect. doi: 10.2139/ssrn.2401257.

Simon, H. A. (1967). Motivational and emotional controls of cognition. *Psychological Review, 74*(1), 29.

Simons, D. J., & Chabris, C. F. (2012). Common (Mis)Beliefs about Memory: A Replication and Comparison of Telephone and Mechanical Turk Survey Methods. *PloS One, 7*(12), e51876. doi: 10.1371/journal.pone.0051876

Slovic, P. (1972). *From Shakespeare to Simon: Speculations--and some evidence-- about man's ability to process information*. Oregon Research Institute Monograph, (2, 12). Eugene, Oregon.

Spencer, A. (2003). The TTO method and procedural invariance. *Health Economics*, 655-688.

Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior research methods, 43*(1), 155-167. doi: 10.3758/s13428-010-0039-7

Stalmeier, P. F., Wakker, P. P., & Bezembinder, T. G. (1997). Preference reversals: Violations of unidimensional procedure invariance. *Journal of Experimental Psychology: Human Perception and Performance, 23*(4), 1196-1205.

Stiggelbout, A. M., Eijkemans, M. J., Kiebert, G. M., Kievit, J., Leer, J. W., & De Haes, H. J. (1996). The 'utility' of the visual analog scale in medical decision making and technology assessment. Is it an alternative to the time trade-off? *Int J Technol Assess Health Care, 12*(2), 291-298.

Szafran, R. F. (2011). *Answering Questions With Statistics*: SAGE Publications.

Ternent, L., & Tsuchiya, A. (2013). A note on the expected biases in conventional iterative health state valuation protocols. *Medical Decision Making, 33*(4), 544-546.

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

Unic, I., Stalmeier, P. F., Verhoef, L. C., & van Daal, W. A. (1998). Assessment of the time-tradeoff values for prophylactic mastectomy of women with a suspected genetic predisposition to breast cancer. *Med Decis Making, 18*(3), 268-277.

van der Pol, M., & Shiell, A. (2007). Extrinsic Goals and Time Trade-Off. *Medical*

Decision Making, July-Aug, 406-413.

van Nooten, F., & Brouwer, B. (2004). The influence of subjective expectations about length and quality of life on time trade-off answers. *Health Economics, 13*, 819-823.

van Nooten, F., Koolman, X., Busschbach, J., & Brouwer, W. (2014). Thirty down, only ten to go?! Awareness and influence of a 10-year time frame in TTO. *Quality of Life Research, 23*(2), 377-384.

van Nooten, F. E., van Exel, N. J. A., Eriksson, D., & Brouwer, W. B. F. (2016). “Back to the future”: Influence of beliefs regarding the future on TTO answers. *Health and Quality of Life Outcomes, 14*(1), 1-7. doi: 10.1186/s12955-015-0402-6

van Nooten, F. E., van Exel, N. J. A., Koolman, X., & Brouwer, W. B. F. (2015). “Married with children” the influence of significant others in TTO exercises. *Health and Quality of Life Outcomes, 13*, 94. doi: 10.1186/s12955-015-0276-7

van Osch, S. M. (2007). *The Construction of Health State Utilities.* (PhD), Leiden University Medical Center (LUMC), Leiden University.

van de Wetering, E.J., van Exel, N.J.A., Brouwer. W.B.F. (2016). Health or Happiness? A Note on Trading Off Health and Happiness in Rationing Decisions. *Value in Health* 19, 552-557.

End of Chapter 2 Notes

i. IP addresses serve as unique identifiers for Internet locations. Duplicate addresses were avoided in order to minimize the chances that a single respondent would repeatedly complete the survey.

ii. There is no defined interval that is generally accepted within which CPTO is considered to hold. Craig and Busschbach (2009) and Hutchins et al. (2015) reported utility values using proportions with 95% confidence intervals.

Chapter 3
When the Timing is Right:
How Duration Influences the Time-Trade-Off
for Older and Younger Respondents

3.1 Abstract

This study aimed to investigate individual-level trade-offs between length and quality of life using iterations of the time trade-off (TTO), a decision task where respondents are asked to state the number of years they would be willing to give up from a certain fixed time period in a deteriorated health state in order to live in full health. When trading-off longevity for quality of life, respondents have been shown to be particularly hesitant to give up longevity since they are incurring a *loss* on this dimension. Whether respondents of different ages perceive losses differently as a function of their own life expectancy and the time horizon offered in the TTO is a topic on which existing evidence has been inconclusive. This study aimed to clarify these relationships through a within-subjects investigation across several TTO time horizons and complementary qualitative interviews.

TTO values were elicited from respondents categorized into an older or a younger age group. In standard TTO elicitations, constructed with typical, relatively short, time horizons (10-25 years). It was hypothesized that, compared to older respondents, younger respondents would be less willing to give up length of life. In a TTO where the time horizon was the respondent's subjective life expectancy, it was hypothesized that younger respondents would provide lower health state values than older respondents due to decreasing marginal utility for life years.

The hypothesis that younger respondents would exhibit greater reluctance to trade length of life for HRQoL was not supported, as a significant difference was only observed between young and older valuations in standard TTOs in the 12-year TTO for Crohn's disease (0.63 for younger respondents and 0.83 for older respondents). Moreover, the difference was opposite to that predicted – younger respondents were *more* willing to forgo length of life for improved quality of life than were older respondents. No significant differences between age groups emerged for the Crohn's disease 25-year TTO, the migraine 10-year TTO, and the back pain 15-year TTO. The observed effect in the 12-year TTO for Crohn's disease may be due to an interaction between the health state and time horizon (which differs by age), or, due to chance event and thus would warrant

confirmation in future studies.

The hypothesis that older respondents would have higher values in LE TTOs was also not supported, as the LE TTO values for Crohn's disease were lower in older respondents, though not significantly and no differences were observed in LE TTOs for depression, the other health state evaluated using the LE frame. Notably, more than a quarter of respondents were unwilling to trade in the LE TTO for Crohn's disease.

There was some evidence of proportional decision strategies which emerged in quantitative and qualitative data. Future studies may wish to investigate the circumstances in which respondents adopt particular decision strategies and the degree to which factors such as familiarity with the health state and the search procedure affect these strategies. This study also presents several reasons to consider and barriers to LE TTO, particularly in terms of understanding how endowing respondents with their full life expectancy influences trade-off behavior.

Overall, the results highlight that the TTO yields inconsistent values from respondents of different ages and that these values are significantly affected by the time horizon used in the task. These findings expand the existing body of work on systematic violations of fundamental TTO parameters. Further, these findings raise important questions about which values should be used in health care prioritization.

3.2 Background

In health care, as in all public sectors, resources are limited and therefore priorities must be set. Currently, across several national health care systems, health care resources are prioritized with the help of a particular type of cost-effectiveness analysis – cost-utility analysis – that is based in part on a composite measure of disease burden, the quality adjusted life year (QALY). The QALY measure involves both quality (specifically, health-related quality of life, hereon HRQoL) and length of life considerations and enables comparisons between different outcomes associated with different illnesses and disabilities.

A fundamental challenge in eliciting health state values (i.e. the quantification of HRQoL) for input into QALY calculations is ensuring that the values mean what they should (i.e. they represent quality of life in a given health state, the *validity* of the construct) and that they are reproducible and consistent (the *reliability* of the construct). The reliability and validity of health state values have been brought into question by the observation of systematic violations of the assumptions from which the QALY is derived.

A growing body of research has shown that values are affected by attributes of the respondents from whom they are elicited and by how they are elicited.

The time trade-off (hereon TTO) is a commonly used choice-based method for eliciting health state values for implementation into QALY calculations. In its standard form, the TTO asks respondents to state the number of years they would be willing to give up from a certain fixed time period in a deteriorated health state in order to live in full health. For example, a respondent may be asked how many years they would be willing to forgo to live in full health as opposed to living in health state *x* for ten years. A TTO value is calculated from the number of years the respondent requires in full health divided by the time horizon (e.g. ten years in the example just given). The following investigation has the aim of better understanding how construction of the TTO (specifically, the length of the time horizon) and the attributes of respondents (younger or older) affect the health state values that are elicited.

Given the time dimension encompassed in the TTO, an important consideration is the age of the respondent (and therefore, specifically, their perception of the time horizon). In the following sections, the concept of loss aversion is reviewed along with discussion of the debate surrounding whose preferences should be used in heath state valuation. Subsequently, possible causes of age differences and the existing literature on age differences are presented followed by an introduction to the LE TTO.

3.21 Loss Aversion in the TTO

The notion of having to *give up* life years has been identified as a possible source of inconsistency in health state values as respondents, both in empirical and real-world settings, across numerous domains, have been observed to be loss averse. Loss aversion implies that individuals asymmetrically value gains and losses such that if they are asked to forgo something in one instance (e.g. sell *x*) and to accept an equal outcome in another (e.g. buy *x*), their valuations (i.e. *selling* and *buying* values) would be inconsistent (such that $x_{gain}<x_{loss}$, the *buying* price would be outweighed by the *selling* price). This phenomenon implies that outcomes are not evaluated as absolutes (i.e. where the value of *x* is dependent on whether it is being acquired or given up), but rather as gains and losses corresponding to a reference point (Tversky and Kahneman, 1991).

Looking at the standard TTO, if losses loom larger than gains, there will be an aversion to trading (losing) life years and TTO values will be upwardly biased. Suppose a respondent is presented with (Q_1, T_1) (seen graphically in Figure 1-5 in the Introduction), which (under TTO model assumptions) constitutes their reference point, the location from which gains and losses are evaluated. The respondent is asked to trade-off a gain in health status (from Q_1 to full health, FH) against a loss in terms of duration (T_1 to, for example, T_2^1). The resulting T_2^1/T_1 ratio is the TTO value (note that T_2 may fall anywhere between the intersection of the axis and T_1, however, in the example given here, T_2^1 is used). In line with the explanation provided by Bleichrodt (2002), in the absence of loss aversion, T_2^1 is the number of years indicated by the respondent, whereby the gain in utility resulting from an increase in health status is equal to the loss of utility in terms of life years. If we are to assume that the respondent is loss averse, they will weigh the loss in terms of time (T_1-T_2^1) more than the gain in quality of life (Q_1-FH) and will therefore prefer the initial starting point (Q_1, T_1) over (FH, T_2^1). T_2^1 would need to increase to T_2^2 in order for the respondent to be indifferent between the two outcomes (i.e. Q_1, $T_1 \approx$ FH, T_2^2). As a result, loss aversion will have an upward bias on TTO values since T_2^2/T_1 will exceed T_2^1/T_1.[i] For instance, imagine the respondent is asked to trade-off length of life from a given health state for ten years in order to live in full health. In the absence of loss aversion, they may indicate that they would consider living 7 years in full health as equivalent to living ten years in the given health state (value = 7/10 or 0.70); however, if the respondent is loss averse this value will be upwardly adjusted or inflated to, say, 8 years (value = 8/10 or 0.80).

3.22 Whose Preferences?

A particular source of contention in the derivation of health state values has been *who* are the most appropriate subjects from whom to elicit values. ‘Whose values should count’ is a question that has been repeatedly raised among policy-makers, economists and philosophers[ii] with much of the debate surrounding empirically observed differences in values elicited from patient samples and from the general population (Stamuli, 2011).

It is possible that treating the general population as a subgroup in and of itself is a reductionist strategy, clouding significant variability in individual preferences, and importantly, perhaps systematically by subgroup. It is possible that different complex processes and rationales underlie subgroups’ decisions to such a seemingly simple question (TTO) and that more empirical work should be dedicated towards investigating inconsistencies within this group. As is the norm in this type of research, many of the methodologically-oriented studies of health state valuation tasks have used convenience samples, often composed of (relatively young) students. Although these samples provide useful grounds for testing certain behavioural patterns, different general population subgroups may have markedly varying preferences for reasons which have yet to be firmly identified (e.g. systematically different use of response strategies/decision processes by age group). The addition of other age groups whose values can be compared to those of younger respondents may help to identify biases, such as loss aversion, on health state values.

It would seem unsurprising that a particular number of years or length of remaining life holds meaningful differences between age groups, and therefore that respondents of different ages may provide varying health state values (see Chen et al., 2011 and Ma, 2010 for comprehensive reviews on age-related decision-making). Drawing from concepts rooted in psychology, economics and the general decision-making literature, three broad hypotheses for an age-based variability in health state values are set out in Box 3-1. Note the varying directional influence on values of each hypothesis.

Box 3-1: Overview of possible causes of age differences in health state values

Emotional regulation: Emotional regulation is at the core of socioemotional selectivity theory, which posits that with a decreasing time horizon individuals become increasingly more selective in terms of the information to which they pay attention (Cartensen, 1993; Li et al., 2013). Older adults have been observed to be relatively more skilled at disallowing negative emotions to override situations or decisions, as evidenced in research by Carstensen et al. (2000) (see also Blanchard-Fields et al., 1995; Isaacowitz et al. 2008; Mather and Carstensen, 2005). This may translate into a decreased willingness to trade-off from a poorer health state in the TTO, resulting in higher health state values, if the older respondent is able to anticipate coping with its negative aspects.

Value creation: With a similar influence as emotional regulation, Johnson et al. (2007) hypothesize that age may mediate value creation (i.e. preferences) formed through the interaction of memory, knowledge, and experience. Whilst deriving relative value or preferences for one outcome over another, individuals may rely on what they know about comparable products or similar experiences (Lanteri and Callabelli, 2007). In terms of the TTO, it may be that older respondents have more experience or a greater familiarity with a particular health state, increasing their anticipated ability to cope and rendering them less willing to trade longevity for HRQoL in the TTO. This would result in increased health state values. Conversely, it is also possible that older respondents would be more willing to trade-off duration for HRQoL if, due to greater life experience with similar situations, they perceive the burden of illness to be high. For example, it has been shown that as the severity of the impaired health state increases, older respondents provide lower health state values than younger respondents (Arnesen and Trommald, 2005; Bleichrodt and Pinto, 2002; Dolan et al., 1996; Lundberg et al., 1999). They may be less loss averse, i.e. have a greater willingness to forgo duration for HRQoL improvements than younger respondents, and thus provide lower health state values.

Prospect theory: Prospect theory was developed (originally in the context of monetary outcomes) to accommodate observed behavioral deviations from standard economic theory for which a key tenet is the notion that individuals have stable, well-defined preferences. Two major elements of prospect theory, loss aversion and diminishing marginal utility for additional benefit (e.g. each incremental year holds progressively less worth), may have varying influences on values of younger and older respondents. With regards to loss aversion, older respondents may not perceive standard TTO time horizons (T_1) (i.e. 10 or 20 years) as being as much of a threat to their life expectancy compared to younger respondents. For younger respondents, this conflict between anticipated life expectancy and the time horizon used in the standard TTO is predicted to increase loss aversion and thus increase the health state values they provide. In terms of the influence of marginal utility (i.e. *discounting* of future life years), if trade-offs are occurring at relatively long time horizons, then later years will be discounted at an increasing rate, thus traded-off more easily, resulting in lower TTO values.

Beyond the three concepts presented in Box 3-1, a number of age-related factors remain to be explored in the literature. For example, cohort or generational effects may affect expectations for longevity and/or health. Older generations, for example, having lived parts of their lives under different societal circumstances (e.g. the Depression and WWII) may perceive length or HRQoL in a different manner than younger generations who have not experienced comparable events. Another source of age-difference could involve variations in the expectation to benefit from future innovative health interventions. Despite these possibilities among many others, in general, there is a significant lack of qualitative findings to substantiate observed age differences in the literature (and therefore there is little reason to suspect that inconsistencies are caused by cohort effects or, age-related information processing capabilities).

Kovalchik et al. (2005, p. 80) remark that “[it] is conceivable that our scientific model of [economic] decision making, so heavily rooted in studies of 20-year-old students, is a misleading guide to the behavior of older people” (for similar arguments see also Johnson et al., 2007, Panidi, 2012 and Sherbourne et al., 1999). From a social policy standpoint, if health state values are dependent on the age of the respondents, an alternative use of resources may be advocated for based on sample characteristics. This may be seen as an analogous argument to the patient-general population debate, both of which create critical challenges in determining whose values should be used in health state valuation (Oliver and Wolff, 2014).

3.23 Interactions Between the TTO Time Horizon and Respondent Age

The TTO is commonly constructed using relatively short time horizons of 10 or 20 years (Arnesen and Trommald, 2005). A key assumption that underlies the TTO, and that must be met in order for the TTO to comply with the parameters of economic theory, is that respondents hold the same utility for time regardless of the time horizon, termed constant proportional trade-off (CPTO). CPTO implies that the relative value of each individual year in a 40-year TTO, for example, will be the same in a 10-year TTO so that if the respondent were to indicate indifference at 20 years in full health in the 40-year TTO for health state *x* (value = 20/40 or 0.50), they would (proportionally) opt for 5 years in the 10-year TTO for health state *x* (value = 5/10 or 0.50).

There are mixed views in the literature as to whether CPTO holds, with some support (Bleichrodt and Johannesson, 1997; Dolan and Stalmeier, 2003; Pliskin et al., 1980; Stalmeier et al., 1997) as well as contrasting evidence (e.g. see Attema and Brouwer, 2010; Lin et al., 2012; Sackett and Torrance, 1978; Stiggelbout et al., 1995). This study proposes that CPTO may not hold if TTO values are a function of the *remaining years* the respondent expects to live (calculated based on their subjective life expectancy (SLE) – the age to which they expect to live – from which their current age is subtracted).[iii] This hypothesis is founded in the concept of loss aversion and the rationale is as follows.

If the time horizon (T_1) used in the TTO is shorter than the remaining years the respondent anticipates living, they may feel that they have been short-changed in terms of how long they have been given to live (even if they remain in poor health for T_1) (Dolan et al., 1996). Consequently, respondents may be reluctant (i.e. loss averse) to trade-off from the time horizon given in the TTO since the time horizon itself already constitutes an abbreviation of their expected remaining years. Suppose, for example, that the respondent anticipates living 45 more years until their death and the time horizon used in the TTO is ten years. Even if the respondent were unwilling to trade from the 10-year time horizon, this would still constitute a loss of 35 years from their initial expectation for the SLE.

3.24 Life Expectancy TTO

An alternative form of the TTO that attempts to resolve the issue of inconsistency between one's remaining years and the time horizon employs a duration that approximates the respondent's actuarial or personal SLE (hereon termed the LE TTO). Respondents are asked to state their SLE or it may be taken from available demographic data (e.g. population averages) or life tables. In this study, the variable 'remaining years' served as the time horizon for the LE TTO (as above, calculated by subtracting the respondent's age from their SLE).

It is possible that a standard TTO (assuming a typical 10-year time horizon) and the LE TTO could be similar constructs for older respondents. For example, suppose the respondent is 65 years old and the TTO time horizon is 10-15 years. For this respondent, a standard time horizon would coincide with their remaining years if they expected to live until 75-80 years. In contrast, for younger respondents, a time horizon of 10-15 years likely represents a substantial loss from their expected number of remaining years. Note that in the LE TTO, however, younger respondents will be considering trades from a longer expected period of remaining life than older respondents and thus may be influenced to greater extent by decreasing marginal utility (as referred to in Box 3-1 under prospect theory) which would be expected to downwardly influence values.

The use of a respondent's remaining life expectancy as the time horizon has received limited attention in research on QALY methodology. A number of clinically-oriented studies, however, have used LE TTOs to evaluate condition-specific quality of life including arterial disease (van Wijck et al., 1998), breast hypertrophy (Chang et al., 2001), visual acuity (Sharma et al., 2002), ulcerative colitis (Waljee et al., 2011), rheumatoid arthritis (Buitinga et al., 2012), and hypoglycemic events (Evans et al., 2013), for example. However, given that these studies were not aimed at evaluating the methodological rigor of the TTO, they largely omitted TTO values generated from different sample groups (e.g. respondents of different ages) and TTOs constructed with other time horizons from both of which the underlying assumptions of the TTO could be tested.

3.25 Other Studies

Some authors have postulated that older respondents allocate greater weight to quality of life than younger respondents. This may be because they have lower expectations for their duration of their remaining life, and/or, they are more willing to give up years as they anticipate a rapid decline in quality of life towards the end of the time horizon (Sommers et al., 2008).

If age differences emerge in health state valuations, they may result from legitimate differences in the perception of the health state (e.g. due to the influence of their own current health, assuming it differs across ages) (Essink-Bot et al., 2007). Respondents of different ages may also hold varying time preference rates – for example, if the time horizon is greater or inferior to the respondents' expectation for longevity (Kind and Dolan, 1995).

Whether systematic variation by age exists in health state valuations remains unresolved in the literature and the findings for an effect of age on health state valuation are mixed. Augestad et al. (2013), Best et al. (2010), Souchek et al. (2005), Fisman (2005), and Hsu et al. (2012) found a positive relationship, while Ayalon and King-Kallimanis (2010), Shimizu et al. (2008), Lundberg et al. (1999), and Zarate et al. (2008) found that age and health state values were inversely correlated. Carter et al. (1976) and Rosser & Kind (1978) found no evidence that health state valuations correlated with age, sex or socioeconomic status. Further, Sackett & Torrance (1978), for instance, found an age effect for some but not all health state valuations with the TTO.

Several studies have compared trade-offs at different time horizons, some using LE TTOs. They have primarily been tests of the assumption of CPTO or have drawn samples from patient populations (which tend to have limited age ranges). Thus, most studies have not purposefully sought to define the role of SLE and remaining years at different time horizons and lack the necessary data points to infer that trade-offs might be a function of SLE or remaining years. It is nonetheless useful to review these studies since their findings may offer insight into the results of this study.

Using a sample of pertussis patients ranging in age from 18-84 years, Lee et al. (2005) conducted an eight-week TTO and a LE TTO and witnessed no significant differences in values. Similarly, Tosteson et al. (2002) looked at a LE and a one-year TTO for benefits of pharmacological agents (n=27, 25-63 years of age), finding similar values between the two iterations. Attema and Brouwer (2012) had a sample of students evaluate ten time horizons ranging from one year to an actuarial life expectancy and also found no clear relationship between TTO values and duration. Stiggelbout et al. (1995) hypothesized that colorectal cancer patients with longer actuarial life expectancies would be more willing to trade-off duration for HRQoL than patients with shorter expected survival; however, no such relationship was found. In addition, some respondents found the time horizons they were presented with to be unrealistically long.

Other studies have found more discernible patterns between the TTO time horizon and values. In an early study of health state valuations, Sackett and Torrance (1978) had members of the general population of all ages value a range of conditions across three time horizons (3 months, 8 years, and the remainder of their lives) and found that as the time horizon increased, respondents' values decreased. Essink-Bot et al. (2007) evaluated three groups of respondents at different life stages (20–25 years, 45–50 years, 60–65 years) using two variants of the TTO, a 10-year TTO and an actuarial LE TTO. They found that the LE TTO produced lower values than the 10-year TTO, meaning that respondents were more willing to give up life years at longer durations. There was no significant effect of age on TTO values; notably, they did not assess the relationships between respondents' SLE and their TTO values. The findings of these studies are consistent with the proposition that at longer durations, duration and HRQoL are increasingly interchangeable, which may be due to a decrease in conflict between their SLE and the time horizon in the tasks and/or decreasing marginal utility.

Similarly, using predefined proportions (e.g. 75% to 80% or 20% to 25%) of the respondents' actuarial remaining life expectancy, Heintz et al. (2013) found that, for respondents who expected to live longer than the time horizon, the greater the difference between their remaining life expectancy and the TTO time horizon, the less willing they were to trade longevity for better health. Likewise, respondents with remaining LEs shorter than the time horizon were increasingly willing to trade off duration for quality of life, and as the difference between their remaining LE and the time horizon became larger (SLE < time horizon), they were increasingly likely to make trade-offs. Van Nooten and Brouwer (2004) witnessed a positive relationship between SLE and TTO scores; however, they did not investigate a shorter time horizon than actuarial life expectancy. They observed that when using an 80-year TTO, respondents with higher SLEs were less willing to forgo duration to improve HRQoL. In a follow-up study, van Nooten and Brouwer (2009) found that when respondents were asked to complete a series of 10-year TTOs, they were less likely to trade longevity for an improvement in HRQoL as their SLE increased.

To date, the interaction between the TTO time horizon and age has been investigated for the most part incidentally, with studies of either CPTO or age effects in isolation being more common (e.g. Dolan and Roberts, 2002). Studies that have focused on the interaction between age and the time horizon used in the TTO have yielded inconclusive results. This is an important topic from both a policy perspective (e.g. given age groups vary significantly in their consumption of health care goods and use of services and the corresponding contentious *whose values* argument) as well as in assessing the methodological strength of the TTO instrument.

3.3 The Study

Using both standard (defined in the context of this study as TTOs with shorter time horizons, specifically, less than 25 years), and LE TTOs, this study aimed to investigate how TTO values are affected by the structure of the TTO task (specifically, the time horizon) and the respondent's age. A main objective of the study was to assess whether trade-off behavior was a function of the time horizon and respondents' remaining years. This study also served as an exploratory investigation of the LE TTO using two different age groups. Further, unstructured qualitative data were elicited to shed light on why respondents were willing or unwilling to make trade-offs and whether decision rationales varied systematically by age or time horizon.

Based on what is known about the dynamics of loss aversion, age, and TTO valuations from previous studies as well as from limited empirical assessments of the LE TTO, the following propositions were put forward:

Propositions:

1. *Standard TTO values for younger respondents (YA) will be higher than standard TTO values for older respondents (OA) (due to greater loss aversion in younger respondents, stemming from conflict between remaining years and the time horizon)*

$$vTTO_{standard*YA} > vTTO_{standard*OA}$$

2. *LE TTO values for older respondents will be higher than LE TTO values for younger respondents (due to diminishing marginal value assigned to distant years in the longer LE time horizons for younger respondents)*

$$vTTO_{LE*OA} > vTTO_{LE*YA}$$

3.4 Methods

Prior to beginning the elicitation exercises, respondents completed a number of questions providing information on their age, gender, SLE, education and perceived current health. A comprehensive body of research has indicated that certain demographic variables influence health state values (e.g. see Dolan and Roberts, 2002, or Wittenburg and Prosser, 2011).

Eighty respondents were recruited through the Behavioural Research Laboratory (BRL) at the London School of Economics (LSE). Respondents received remuneration of £5 for their participation. The respondent pool available from the BRL consists largely of higher education students of various nationalities although it is accessible without restrictions to anyone over the age of 18 (e.g. LSE students, students from other universities, LSE staff members, individuals working in the local area or who are regular lab participants in paid studies across London).

<table>
<tr><th>Respondent Groups</th><th>TTO tasks</th></tr>
<tr><td>Younger respondents (<40 years old)</td><td rowspan="2">- 10-year, migraine
- 12-year, Crohn’s disease
- 15-year, back pain
- 25-year, Crohn’s disease
- LE TTO Crohn’s disease
- LE TTO depression</td></tr>
<tr><td>Older respondents (>40 years old)</td></tr>
</table>

Figure 3 - 1: Study design

Questions were presented to respondents randomly to prevent ordering effects. Three health states were evaluated using the standard TTO: migraine, Crohn’s disease and back pain. The selection of health states and time horizons was intended to facilitate comparisons with another study not included in this thesis and with other published studies. Moreover, depression was evaluated using an LE TTO framework since a longer time horizon seemed to be a realistic option for this mental health condition. Time horizons were selected for the standard TTOs on the basis that they would represent realistic expectations for longevity in both age groups (i.e. respondents in both groups could imagine living at least *x* more years). It is also the case that the standard TTO time horizons selected are similar to those commonly used, increasing the generalizability of the results. Generally similar but not identical time horizons were used across health states to encourage the respondent to carefully consider both the duration and the quality of life associated with the given health state (e.g. by rendering it more challenging to abide by a consistent heuristic for each trade). Migraine and back pain TTOs were presented to respondents with time horizons of 10 and 15 years, respectively. In order to gain a closer look at whether there is a discernible interaction between the time horizon, health state values, and age, Crohn's disease was assessed using two standard TTOs: a 12- and a 25-year version.

Two LE TTO elicitations were included, one valuing Crohn's disease and a second, looking at depression. The depression health state was selected for explorative purposes to evaluate how older and younger respondents perceive a predominantly mental health condition as opposed to the relatively more physically debilitating states included in the questionnaire. The respondent's age was subtracted from their subjective life expectancy to calculate the LE time horizon. Implementing an actuarial life expectancy would leave open the possibility that the time horizon would still appear to be a loss if the respondent is optimistic about his or her life expectancy (Kattan et al., 2001). In behavioural economic terms, if the respondent's SLE does not align with the life expectancy provided in the TTO, there is a potential mismatch between the reference point provided in the task (i.e. the time horizon) and the reference point from which the respondent gauges their trade-offs (i.e. their SLE). It would then follow that actuarial LE might be perceived as a loss from SLE, pushing health state value upward.

The search procedure used to elicit values has been shown to affect the values in different directions depending on the technique which is used. The short duration (10, 12, 15, and 25-year) TTOs were elicited via an upward titration technique. In an upward titration search procedure, the respondent selects between a number of options presented to them until they reach the point at which they are indifferent between the two options (e.g. 1 year in full health or 10 years in health state *x*, 2 or 10, 3 or 10…). Based on results from earlier studies, a 'fill-in-the-blank' technique (where respondents are not given options to choose from and are instead required to state their value) and a downward titration technique (e.g. 9 years in full health or 10 years in health state *x*, 8 or 10, 7 or 10…) were avoided in the shorter duration TTOs. These techniques have been shown to be particularly vulnerable to loss aversion (i.e. shorter time horizons may be interpreted as significant losses from the respondent's SLE and these techniques compound the perception of this loss) (Attema and Brouwer, 2012; Tversky and Kahneman, 1991). Dolan (2011), for instance, observed that downward titration and 'fill-in-the-blank' techniques yield higher values, potentially indicative of a stronger influence of loss aversion. Moreover, it was thought that an upward titration search procedure would be an effective way to elicit the respondent's absolute lowest threshold for trading-off length for HRQoL (Dolan and Stalmeier, 2003).

In the LE TTO, either the upward or downward titration method would involve a cumbersome number of iterative steps in order to arrive at the respondent's valuation, which might decrease engagement in respondents. Although several studies have used a bisection procedure to elicit LE TTO values (Essink-Bot et al., 2007; Heintz et al., 2013; Sackett and Torrance, 1978), in a pilot study, many respondents indicated that they decide on an acceptable range or value after reading through the question, prior to the elicitation task itself. A 'fill-in-the-blank' procedure, similar to that used in Van Nooten and Brouwer (2004) was considered the most suitable and the LE TTOs in this study were elicited in this manner.

3.41 Analyses

Quantitative analyses were carried out in SPSS v21. Table 3-2 outlines mean and median values for each health state. The majority of the analyses focused on median values since they are relatively unaffected by outliers. Shapiro-Wilk Tests showed that the distributions (except for the 25-year TTO for Crohn's disease in the younger group and the migraine and back pain TTOs in the older group) deviated significantly from normality and as a result nonparametric tests were used. At an aggregate level, Mann-Whitney U tests (which assess the null hypothesis that both groups in the study have the same median) and Wilcoxon Rank tests (a repeated measures design which assesses whether values provided by respondents in the same group, i.e. younger or older, differ across conditions, or time horizons) were conducted. Mann-Whitney U tests were used to compare median values between younger and older respondents – e.g. comparing median 10-year TTO values for migraine between younger and older groups. Wilcoxon Rank tests were used to make within group comparisons – e.g. how do the median 12- and 25-year TTO values for Crohn's disease differ for younger and older age groups, respectively. The distributions of values for each health state and duration were plotted by age group.

3.5 Results

None of the respondents had any major health problems. It is worth noting that health status might have confounded values insofar as these respondents may have been predominantly older or deviated in their preferences for HRQoL versus longevity. Forty respondents were categorized as *younger* respondents (under 40 years of age) and forty as *older* respondents (over 40 years of age). It should be noted that 90% of the younger respondent group was under 30 years of age. As seen in Table 3-1, the younger respondent group had a median age of 24 years and the older group had a median age of 53 years. Both groups had similar SLEs, gender distributions and approximately half of respondents in each group held an advanced degree. The mean time horizons for the LE TTO (calculated by subtracting mean age from mean SLE) in the young and older groups were 55 years and 32 years, respectively.

Table 3-1: Demographic characteristics

	Younger Respondents	Older Respondents
N	40	40
Sex (female, %)	22 (55)	22 (55)
Age, median (range)	24 (18-39)	53 (40-76)
Subjective life expectancy, SLE (mean)	80	85
Remaining years (mean)	55	32
Education, n (%)		
Secondary school	1 (2.5)	7 (17.5)
Undergraduate degree	17 (42.5)	16 (40)
Advanced degree	22 (55)	17 (42.5)

In terms of demographic variables, there were no significant correlations between respondent age, SLE, remaining years and values elicited at any of the durations except between SLE and the LE TTO values for depression (p=0.049), and age and the 12-year TTO values for Crohn's disease (p=0.023). Educational status and the values elicited in the LE TTO for Crohn's disease were also significantly related (p=0.003).

Table 3-2: Health state values for standard TTOs

	N		Median (IQR)				
	Younger	Older	Younger	Older	Difference	CI*	p value
Crohn's 12-year	40	39	0.63 (0.50-0.73)	0.83 (0.58-1.00)	-0.16	-0.25, -0.08	0.01
Crohn's 25-year	39	36	0.64 (0.44-0.80)	0.80 (0.49-0.96)	-0.08	-0.20, 0.04	0.20
Migraine 10-year	40	39	0.70 (0.40-0.80)	0.60 (0.50-0.80)	0.00	-0.10, 0.10	0.71
Back Pain 15-year	37	38	0.67 (0.52-0.87)	0.67 (0.47-0.80)	0.00	-0.07, 0.14	0.71

IQR = Inter-quartile range
*95% Confidence Interval

3.51 Standard TTO Elicitations

The median standard TTO values (the 12- and 25-year TTOs for Crohn's disease, the 10-year TTO for migraine and the 15-year TTO for back pain) showed no consistent pattern distinguishing between the age groups. Only the median values elicited in the 12-year TTO for Crohn's disease were significantly different between younger and older respondents (p=0.006, Mann-Whitney U test). In this elicitation, older respondents were less willing to forgo length of life for improved quality of life, producing relatively higher median values. Looking *within* each age group, the median values for the 12-year and 25-year Crohn's disease TTOs were similar within the younger and older respondent groups. Thus, the overall findings are not consistent with the hypothesis that values are a function of the difference between remaining years and the time horizon.

Value distributions were plotted in order to gain insight into the relationships between trade-off patterns and age, including trend-lines representing the moving averages for each age group.[iv] As seen in Figures 3-1, 3-2 and 3-4, the respondent groups show comparable value distributions in the standard TTOs for both migraine and back pain health states in younger and older respondents. Figure 3-3, the histogram with the 12-year TTO values for Crohn's disease, appears to be the only instance where there are observable differences between younger and older respondents in the values.

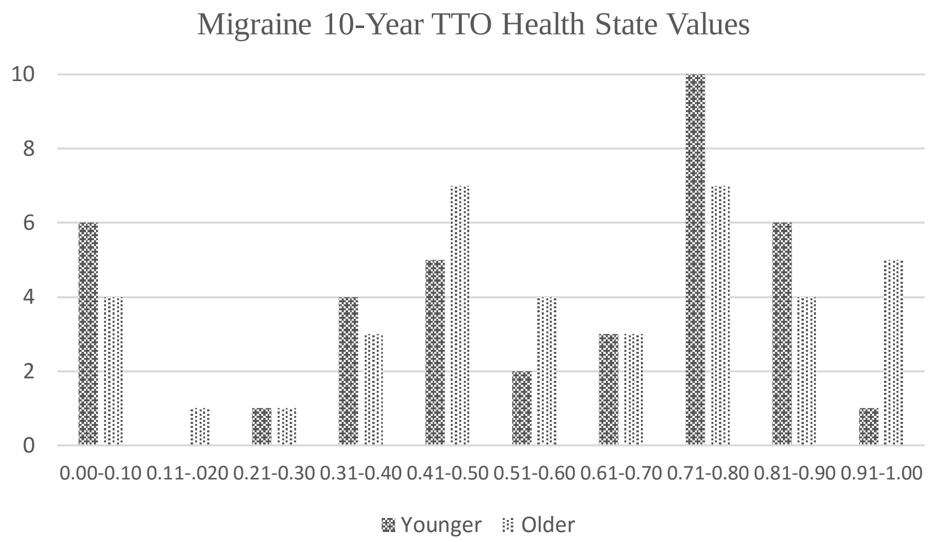


Figure 3-2: Number of respondents with health state values for the 10-year TTO, migraine, in given ranges. Younger respondents, darker shading; older respondents lighter shading.

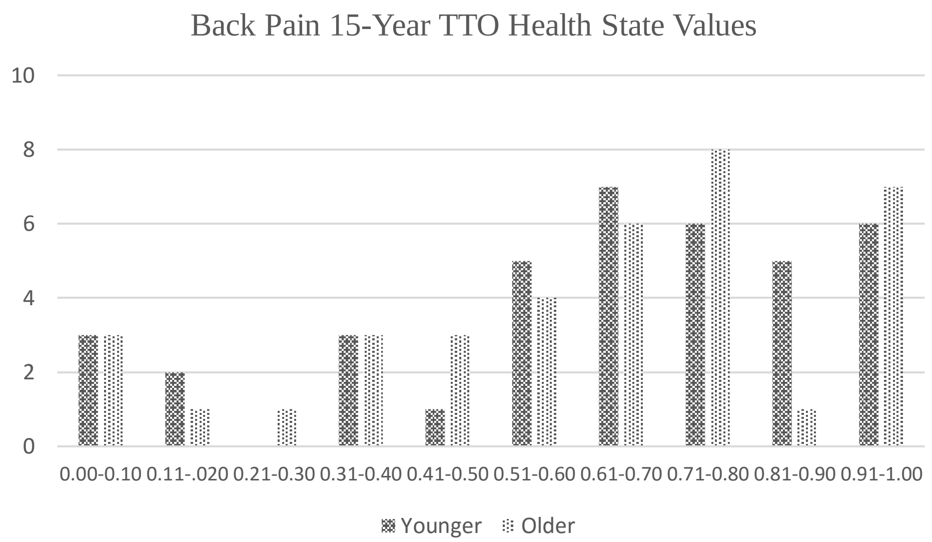


Figure 3-3: Number of respondents with health state values for the 15-year TTO, back pain, in given ranges. Younger respondents, darker shading; older respondents lighter shading.

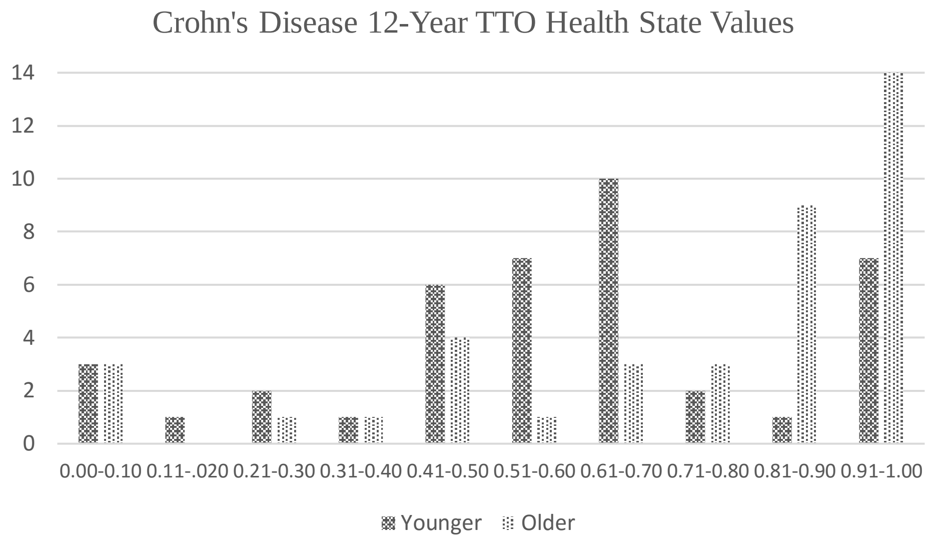


Figure 3-4: Number of respondents with health state values for the 12-year TTO, Crohn's disease, in given ranges. Younger respondents, darker shading; older respondents lighter shading.

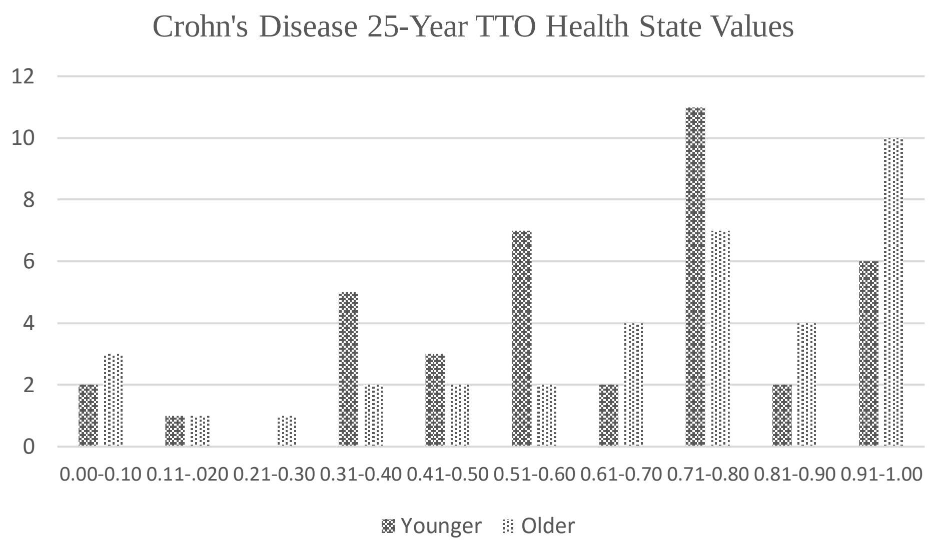


Figure 3-5: Number of respondents with health state values for the 25-year old TTO, Crohn's disease, in given ranges. Younger respondents, darker shading; older respondents lighter shading.

In the 12-year TTO for Crohn's disease (seen in Figure 3-3), a ceiling effect (meaning respondents are unwilling to trade any longevity for health) was observed for older respondents (10 non-traders) whereas almost half of younger respondents (17, 42.5%) cited indifference at 6 or 7 years out of 12.[v]

Qualitative data lent support to the idea that several respondents implemented proportional decision strategies, citing: “I see myself choosing 33%, 8/25 years in health… however, any period shorter than 8 years is too short”, “I wish to trade up to 50% of the entire life span [time horizon]” and “25 years is a long time, I want to live at least 40% of that time.” As seen in Figure 3-1, that 7 older respondents and 5 younger respondents traded-off half of the time horizon in the 10-year TTO for migraine also suggests that respondents may have been tempted to rely on a rule of thumb to arrive at their valuation in this TTO. In the 15-year TTO for back pain, 7 younger and 6 older respondents were willing to forgo one-third of the time horizon, also potentially the result of a proportional strategy.

It was hypothesized that in the standard TTOs, younger respondents would provide higher values than older respondents. However, Table 3-2 shows that younger respondents provided lower median values than older respondents for both 12- and 25-year TTOs, statistically significant for the 12-year TTO.

Comparing the 12-year and 25-year TTOs for Crohn’s disease within age groups, the younger and older groups each provided similar values across the two durations (Figure 3-5). The 12-year and 25-year median TTO values were not significantly different within either age group, although the difference approached significance in the older group (p=0.06).

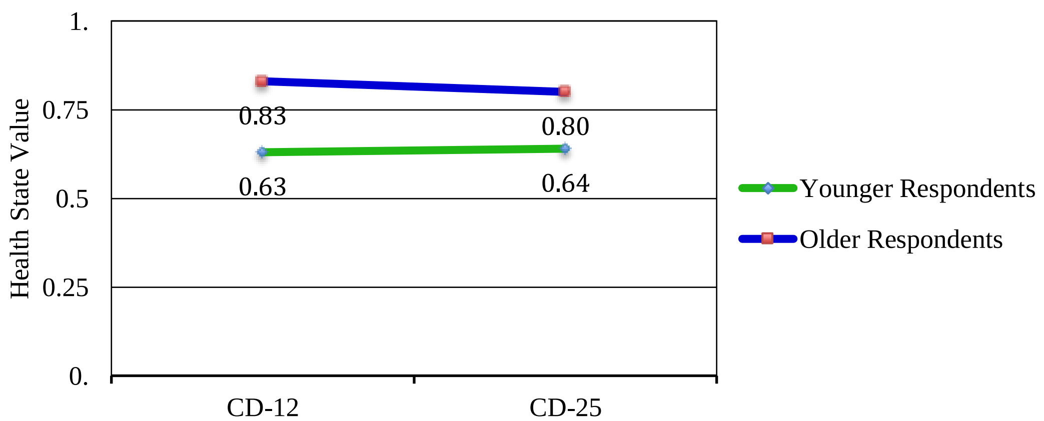


Figure 3-6: Median health state values in younger and older groups in 12- and 25-year TTOs (Crohn's disease)

The direction of the changes in values between the 12- and 25-year Crohn’s disease TTOs at the individual respondent level is shown in Table 3-3. Both higher and lower values at the longer time horizon were commonly observed. One-half of younger respondents allocated a higher value in the 25-year TTO compared to the 12-year.

Table 3-3: Directional change in values with increased time horizon, 12- and 25-year TTOs (Crohn's disease)

	Higher values with increased time horizon	Lower values with increased time horizon	Equal values with increased time horizon	Missing values
Younger Respondents	20	15	4	1
Older Respondents	8	21	6	4

3.52 Life Expectancy TTO Elicitations

The second specific hypothesis evaluated in this study was that older respondents' LE TTO values would be higher than younger respondents' LE TTO values. As seen in Table 3-4, no differences were observed between age groups in either the depression or Crohn's disease TTO. For Crohn's disease, over a quarter of the sample were unwilling to trade (discussed below). Decreasing marginal utility would predict lower values at the longer duration presented in the LE context, especially for younger respondents. It is notable that younger respondents provided higher values for both LE TTOs than they did in any of the standard elicitations.

Table 3-4: Health state values for LE TTOs (Crohn's disease and depression)

	N		Median (IQR)				
	Younger	Older	Younger	Older	Difference	CI*	p value
Crohn's LE TTO	38	39	0.82 (0.64-0.88)	0.76 (0.50-1.00)	0.01	-0.10, 0.13	0.73
Depression LE TTO	38	37	0.84 (0.64-0.92)	0.84 (0.58-1.00)	-0.02	-0.10, 0.07	0.48

IQR = Inter-quartile range
*95% Confidence Interval

There were 4 younger and 12 older non-traders for the Crohn’s disease LE TTO and 6 younger and 17 older non-traders in the depression LE TTO (illustrated in the skewed distribution in Figures 3-6 and 3-7).

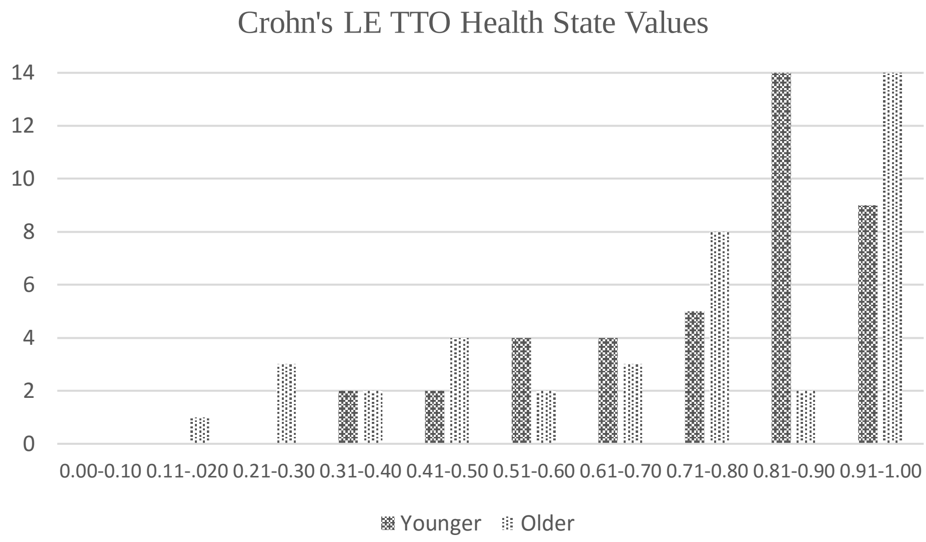


Figure 3-7: Number of respondents with health state values for the LE TTO, Crohn's disease, in given ranges. Younger respondents, darker shading; older respondents lighter shading.

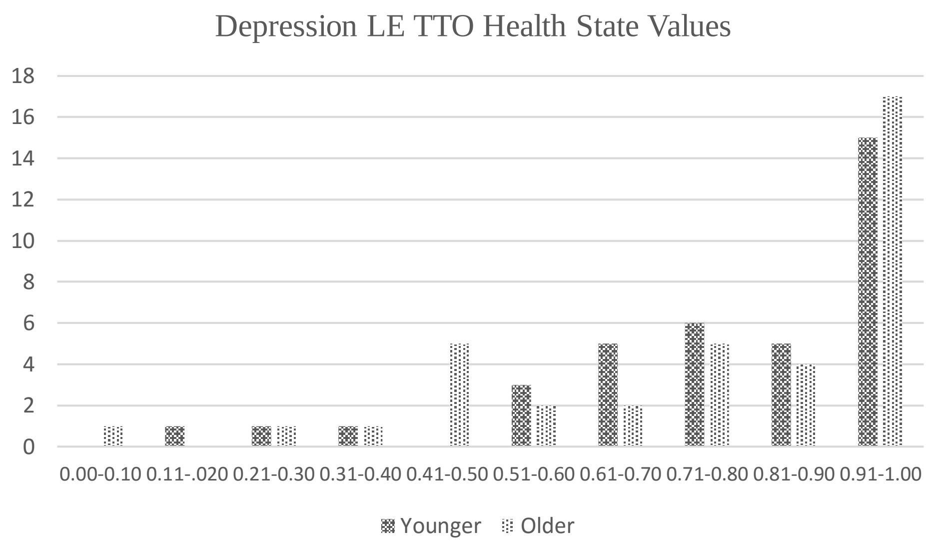


Figure 3-8: Number of respondents with health state values for the LE TTO, depression, in given ranges. Younger respondents, darker shading; older respondents lighter shading.

3.6 Discussion

Age differences in health state valuation merit investigation for several reasons including their importance in relation to cost-effectiveness decisions and resource allocation, and the existing evidence of inconsistencies in health state valuation. This study sought to assess age differences in the standard and LE versions of the TTO, specifically whether the time horizon would affect values based on the degree to which it was perceived as a loss from the number of remaining years the respondents expected to live. It was thought that a difference in values would be seen between younger and older respondents due to loss aversion or diminishing marginal utility depending on whether the time horizons were presented as standard durations or in relation to subjective life expectancy. This hypothesis was refuted across all time horizons. Nonetheless, overall, the results highlight that the way in which the TTO is presented and to whom, can produce highly varying results.

3.61 Standard TTO

The TTO assessing Crohn's disease using a 12-year trade-off yielded significantly different values between younger and older respondents in the opposite direction than predicted (i.e. younger respondents showed lower values). It was striking that there was a significant age group difference in one health state (12-year TTO for Crohn's disease) and not others (15-year TTO for back pain and 10-year TTO for migraine). Two explanations are proposed: either true age differences exist in the perception of different health states or similar decision heuristics are used by all respondents for some health states but not others.

It is possible that due to the artificially brief time horizon in the standard TTO tasks, both older and younger respondents externalize the TTO task to some degree from their own life expectancies and use similar trade-off heuristics. Respondents in both age groups may largely use proportional trade-offs to arrive at their values. Proportional trade-offs imply that respondents are providing some proportion or fraction of the time horizon as their response, as in CPTO. Importantly, the difference between these two concepts (proportional heuristics and CPTO) is that the former is a means of simplifying the decision through heuristic use whereas CPTO implies the respondent does in fact hold the same utility for each incremental year of life.

A number of studies evaluating standard, shorter duration TTOs have contended that respondents use proportional trade-off strategies (e.g. Attema and Brouwer, 2012 and Dolan and Stalmeier, 2003, see below). In addition, wider literature has suggested that cognitive easing strategies, such as reliance on heuristics, may be used when trade-offs are difficult to make as is the case in TTOs with shorter durations (Hogarth, 1991; Slovic, 2002; Abhyankar et al., 2013). Moreover, in health state valuations, respondents may base their decision predominantly on the life-years attribute since this dimension is more easily understood and can be cognitively processed more readily than changes in quality of life. While it is difficult to distinguish between CPTO and decision heuristics, the qualitative data in this study provided evidence in support of the latter.

The 12-year TTO values for Crohn's disease are an exception to the proposition that respondents of both age groups use a proportional heuristic. Among reasons why Crohn's disease might differ from migraine and back pain in terms of decision strategies, older respondents were more likely to have had personal experience or someone they know affected by Crohn's disease (46% versus 22.5% for the younger group). Payne and Bettman (2004) comment that individual respondent characteristics such as prior experience or knowledge, in line with the value creation hypothesis outlined in Box 3-1, may influence whether the respondent uses cognitive shortcuts (i.e. proportional trade-off strategies) when making their decision. Therefore, if younger respondents are relatively less familiar with Crohn's disease (when compared to migraines and back pain) they may be more likely to carefully consider the health state in question as opposed to relying on heuristics to arrive at their decision. Another possibility is that Crohn's disease was perceived to be a less serious condition to older respondents and thus group differences began to emerge. Older respondents may be less bothered by some of the Crohn's disease symptoms (e.g. episodic diarrhoea) and did not consider them worthy of forgoing any longevity.

While several studies have reported higher values as the time horizon increases (e.g. Stalmeier et al., 1997), for the most part respondents provide lower values as the TTO time horizon increases due to decreasing marginal utility for life years (Attema and Brouwer, 2010). Several health care regulatory bodies, including the National Institute for Health and Care Excellence in the UK and the US Panel on Cost-Effectiveness in Health and Medicine, posit that values should be adjusted using a discount rate (usually in the range of 3-6%) (Gold et al., 1996; NICE, 2013). The mixed pattern of trade-off directions in this study poses a challenge to implementing this type of value discount rate, namely the significant proportion of respondents who provided higher than expected values at the longer time horizons, suggesting that they have a negative rather than positive time preference. If respondents discount life years in different directions, as was observed in this study, the assumption that TTO values elicited at different durations can be adjusted for time discounting for through a common discount factor applied to all responses is forcefully undermined. For example, applying a positive (common) discount rate to values provided by a respondent who has negative time preference (who therefore assigns greater value to future than to more proximal outcomes, trading off fewer years at a longer time horizon), would result in even higher values because the positive (common) discount rate would have an upward effect.

3.62 Life Expectancy TTO

In the LE TTOs, no significant age differences emerged, although a large number of older respondents opted not to trade. Interestingly, for both age groups, the LE TTO seems to involve a different set of considerations than do TTOs with shorter time horizons. In particular, respondents seem to focus on either living out their entire life regardless of their health status or achieving particular life goals, a decision strategy that seems to deviate from that used when the time horizon is shorter.

The higher number of non-traders in the LE TTOs compared to shorter duration TTOs lies in contrast to the results of Gyrd-Hansen (2012) and others (e.g. Pliskin et al., 1980, and McNeil et al., 1981) who found that respondents are more likely to trade longevity for quality at longer durations. It may be that the framing of the LE TTO causes older respondents to anticipate that they would be able to cope with the symptoms if it enabled them to live for the *rest of their lives*. It may also follow that when presented with the *rest of their lives*, older respondents allocate more value to these final years if this wording makes them feel pressed for time whereas younger respondents contemplate their ability to cope with the health state for what they perceive to be a rather prolonged period of time. Payne and Bettman (2004) explain that in order to cope with a difficult trade-off, respondents may engage in emotion-focused coping, where they mediate how much and which information they involve in the decision-making process. The most extreme version of this strategy would be to avoid making a decision altogether, remaining with the status quo (their current situation) or opting for a position which is most easily justified (Anderson, 2003; Luce, 1998).

It may be that when respondents are given the option to live through all life stages (as in the LE TTO), this endowment renders them less willing to forgo longevity (compounded by the fact this was likely their expectation outside of the valuation exercise). The apparent change in processing strategy, orienting towards broader life ambitions (as opposed to numerical shortcuts, e.g. proportional heuristics), may be induced by the non-numerical description of the TTO (i.e. qualitative explanation ‘the rest of your life’ instead of a traditional quantitative time horizon such as ten years) may stimulate more qualitative considerations (e.g. those concerning quality of life). An interesting avenue for future research may therefore be to compare decision processes when the time horizon is presented in a quantitative versus qualitative manner.

Hazen (2006) describes goals as “non-status quo reference points” in that they may change perceptions of gains and losses (as they serve as reference points) but may never actually become reality (i.e. the respondent’s status quo). Hazen (p. 4) comments, “like [these] quality-of-life issues, issues related to quality of health can also be recast in terms of goals.” He cites research by Schwartz et al. (2008) who observed that goals including education, health, family, and wealth may affect the value that a respondent assigns to longevity. Loomes and McKenzie (1989) also propose that values may depend on the particular stage in life that the respondent is in and that thresholds may exist that align with various ambitions across the life-course. For example, a respondent who has dependents (life stage) and wishes to see their children get to college (a threshold) may view forgoing length of life for better quality of life differently than would a retired respondent. The use of goals or aspirations as reference points introduces a number of important implications, for instance, the same respondent may adopt a different goal as a reference point depending on the time horizon and thus measure gains and losses in a different way between these two trade-offs. From this perspective, it is easy to see why TTO values would not satisfy underlying model assumptions (e.g. CPTO). Separating potential medical and non-medical influences also poses another challenge to interpreting the validity of resulting health state values. Fowler et al. (1995) remarked that when making trade-offs it can be difficult to separate the willingness to forgo longevity associated with the health state itself versus other possible factors in the respondent’s life, such as having children (Stiggelbout et al., 1995). In their study of the TTO, Mohide et al. (1988) observed that even the health of individuals for whom the respondent takes care of can impact their valuation of a given health state.

The finding that respondents anticipate their ability to cope with depression lies in contrast to several studies’ results which have shown that when attributes of a health state are categorized (e.g. as they are into the EQ5D – a health state indexing system – into mobility, pain/discomfort, usual activities, self-care, and anxiety/depression), anxiety and depression are among the most undesirable characteristics of a health state and incur a strong negative influence on health state values (Burstrom et al., 2014; Versteegh et al., 2013).[vi] There is evidence that experience with the mental health state has a strong influence on values and general population samples have lower rates of maximum endurable time (i.e. where the individual reaches a point in which they no longer positively value longevity in the given health state) than patients (Weyler and Gandjour, 2011). Thus, it appears that the lack of experience of respondents in this study with the health state allowed them to anticipate they would be able to cope, a result congruent with previous studies showing that general population samples provide higher values than patients for depression (Pyne et al., 2009). From a methodological standpoint, it is possible that the number of non-traders in the depression TTO was influenced by the 'fill-in-the-blank' search procedure, which endows respondents with their full life expectancy and therefore may affect their willingness to forgo any duration.

3.7 Limitations

There are limitations to the study. For one, more distinctly different age groups may enable trends to be observed. Another limitation is the relatively small sample sizes. Larger sample sizes would help provide information as to whether robust differences do exist between older and younger respondents, as would assessment of a greater number of time horizons and a broader range of health state severities. Gathering respondents' valuations for the health states in a 'timeless' setting, as in a visual analogue scale, might also offer some insights into age differences in perceptions of health state severity independent of time. In this study, it was thought that considerations related to interview length and efforts to curb decision fatigue among respondents outweighed the potential value of adding additional questions.

Another limitation, particularly relevant to the topic of age, is that expectations for future health were not assessed. Brouwer and van Exel (2005) found that expectations for health were rather low beyond 70 years of age, which rendered respondents more willing to forgo these years. This possibility was addressed in the research presented in Chapter 2.

The LE TTOs were conducted using a 'fill-in-the-blank' procedure since pilot interviews revealed that respondents found this to be the most intuitive format. Assessing health state values across all three durations using a consistent search procedure would allow analyses between the results of all durations; however, given a primary aim was to explore age differences in the LE TTO, ease in task completion was prioritized in this study. Thus, the effects of age were considered through analyses looking separately across standard TTOs and LE TTOs. No research could be found that has assessed the effects of search procedure using an LE (or similar variant) TTO.

Finally, the study was designed to test specific hypotheses with respect to age and TTO values within a behavioural economic framework; however, the findings suggest that different study designs might better reveal differences in how respondents of different ages arrive at their TTO valuations. For example, a think-aloud task would help gain more insight into the considerations factored into TTO decisions according to age, including information as to why or when respondents rely on different decision strategies.

3.8 Broader Implications

In order for the TTO to derive valid and reliable results, a consensus must be reached on two fundamental issues which have been and will undoubtedly continue to be highly debated: the *how* (referring to the construction of the TTO, particularly the time horizon) and the *who* (generally, which respondents).

The *how*, the time horizon used in the TTO will in some situations be dictated by the health state or condition of interest. Health conditions that place significant limitations on life expectancy are intuitively better served by shorter time horizons (Tosteson et al., 2002). It is for the longer-term or chronic conditions for which the TTO was originally developed where the ambiguity lies as to which construction of the TTO (i.e. what time horizon) is most appropriate. There is no optimal time horizon from which values for all conditions can be estimated. Rather, in most circumstances a trade-off must be made between using a realistic time horizon and potential influences such as loss aversion and time preference (Attema et al., 2013).

Questions such as whether respondents should be discouraged from answering heuristically and if thresholds or using broader life goals to gauge trade-offs in the LE TTO is a more favourable approach, must be addressed. Gigerenzer and Gaissmaier (2011) identify a series of questions to help guide further TTO related-research including: Which heuristics do people use in which situations? Are preferences lexicographic for durations below five years? That is, do respondents have a threshold in terms of how long they are given to live, below or above which they focus their decision solely on longevity or quality of life? Based on the findings of the study presented here, at what point do respondents shift from a seemingly quantitative decision strategy at shorter time horizons to considering broader life goals, which appears to occur at somewhat longer time horizons?

Several have argued in favour of using a TTO with a longer time horizon, specifically the LE TTO, for the reason that it is representative of a more realistic scenario to many respondents than shorter time horizons of 10 or 20 years. That one's life expectancy belongs to their endowment (i.e. what the respondent perceives they possess) and that the use of longer time horizons decreases the effect of time preference offers additional support in favour of this construction (Evans et al., 2013; Stiggelbout et al., 1995; Tosteson et al., 2002). It could be argued that having the respondent contemplate both an unfamiliar time horizon and quality of life, as in the standard TTO, is mentally too costly, resulting in the use of simplifying cognitive strategies. Furthermore, it could be reasoned that the LE TTO is appropriate as there are other decision contexts where people are asked to trade-off from their life expectancy, perhaps the most frequently cited example being that of retirement savings (spending now versus spending later).

In contrast, a more widespread use of the LE TTO would meet with challenges such as the comparability of values with those of previous studies and data sets that have implemented shorter time horizons (e.g. the MVH EQ5D tariffs). However, maintaining the standard TTO as a favoured method of elicitation leaves numerous issues to resolve since even in its most basic form, the mechanics of the standard TTO, including influences of the time horizon and age, remain unclear. Additionally, to accept that respondents adopt proportional heuristics when making trade-offs undermines the QALY's attempt to incorporate quality-of-life into economic analysis. The fundamental issue with heuristic problem solving is that attribute non-attendance (in this case, neglecting quality of life considerations) may be viewed as a regression towards basing cost-effectiveness decisions solely on life expectancy improvements.

Alternatively, Devlin et al. (2009) propose that a utility to duration function be established or that duration-dependent sets of values (i.e. a different set of values for different durations) be derived. Others have suggested that a weighting procedure would help to control for variation in individual life expectancies between respondents (Verschuuren, 2006). However, the results of this study suggest that loss aversion may not be accounted for through correction factors alone since different time horizons may elicit different response strategies used by the respondent to arrive at their point of indifference. It would therefore seem unintuitive to try to address loss aversion through a weighting procedure, given that respondents may use one decision strategy at one time horizon (e.g. a compensatory strategy, trading off between longevity and HRQoL) and a different decision strategy at another (e.g. refusing to trade longevity below a particular number of years in order to achieve a particular life goal).

In regards to the *who* should provide health state values, the most often considered comparison involves that between those who have experience with the health state (i.e.

patients) and those without such experience (i.e. the general population, overlooking potentially significant differences *within* each of those groups, such as those based on age. The results of this study show that if the TTO time horizon is short, respondents provide similar values regardless of their age since their decisions may be based on the use of heuristics. Using a longer time horizon intuitively holds different meanings for different respondent groups, depending on their expectations for longevity. Thus, an argument could be made that values should be derived from a sample that has expectations for longevity that are closely aligned with the time horizon used in the trade-off in order to control for effects of discordant SLE and TTO time horizon. In addition, while the focus of this study was on age differences (specifically, the role of the number of remaining years the respondent expects to live), the difference in values between older and younger respondents in the 12-year TTO for Crohn's disease may indicate that experience with illness is a potential factor in which decision strategies are used by respondents. In the 12-year TTO for Crohn's disease, older respondents provided higher values and reported greater experience with this health state. This finding is relevant to the widely debated argument about whether (or when) health state values should be generated from the populations that are faced with these conditions, that is, particular patient groups. Notably, it is also pertinent to the topic of surrogate decision-making[vii], about which, although not explored here, there is a growing body of literature (Bryce et al., 2004; Winter et al., 2003).

It is also argued that those who have contributed through their entire lives to tax-funded health systems should be the ones to provide health state values. Another line of reasoning concerns general public accountability, that is, whether the public should have a role in setting priorities for the health care they are receiving. While there are justifiable reasons for arguments favouring different respondent groups, none has emerged as an overarching, universally appropriate solution. Accordingly, researchers should continue to pursue a better understanding of how the elicitation methods interact with characteristics of different respondent groups and perhaps work towards deriving a set of contextual rules that advocate for the use of certain respondent groups depending on the situation.

If the objective of deriving representative public preferences using the TTO is to be met, a greater dedication towards understanding the interactions between age and health state elicitation instruments is required. Within aging populations, older individuals may use health services with greater frequency and therefore may argue that their preferences should be strongly considered. This study has provided several insights in regards to the consequences of framing the TTO in different perspectives, highlighting the difficulty in predicting how trade-off decisions are made by individuals and within general population subgroups. Understanding subgroup preference patterns and the means of appropriately combining or prioritizing them on a societal level is a contentious but important topic that needs to be addressed in greater depth in future research.

References

Abhyankar P, Volk RJ, Blumenthal-Barby J, Bravo P, Buchholz A, Ozanne E, et al. Balancing the presentation of information and options in patient decision aids: an updated review. *BMC Medical Informatics and Decision Making*. 2013;13(Suppl 2): S6.

Anderson CJ. The psychology of doing nothing: forms of decision avoidance result from reason and emotion. *Psychological bulletin*. 2003;129(1):139.

Arnesen T, Trommald M. Are QALYs based on time trade-off comparable? – A systematic review of TTO methodologies. *Health Economics*. 2005;14(1):39-53.

Attema A, Brouwer B. On the (not so) constant proportional trade-off in TTO. *Quality of Life Research*. 2010;19:489-97.

Attema A, Brouwer WB. Constantly proving the opposite? A test of CPTO using a broad time horizon and correcting for discounting. *Quality of Life Research*. 2012;21(1):25-34.

Attema, A. E., Edelaar-Peeters, Y., Versteegh, M. M., & Stolk, E. A. (2013). Time trade-off: one methodology, different methods. *The European Journal of Health Economics, 14*(Suppl 1), 53-64. doi: 10.1007/s10198-013-0508-x

Augestad LA, Rand-Hendriksen K, Stavem K, Kristiansen IS. (2013) Time trade-off and attitudes toward euthanasia: implications of using ‘death’ as an anchor in health state valuation. *Qual Life Res*, 22(4):705–14.

Ayalon, L., & King-Kallimanis, B. L. (2010). Trading years for perfect health: results from the health and retirement study. *Journal of aging and health*.

Best JH, Garrison LP, Hollingworth W, Ramsey SD, Veenstra DL. (2010) Preference values associated with stage III colon cancer and adjuvant chemotherapy. *Qual Life Res*, 19(3):391–400.

Blanchard-Fields F, Jahnke HC, Camp C. Age differences in problem-solving style: the role of emotional salience. *Psychology and Aging*. 1995;10(2):173.

Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Economics*. 2002;11:447-56.

Bleichrodt H, Johannesson M. Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. *Journal of Health Economics*. 1997;16(2):155-75.

Bleichrodt H, Pinto JL. Loss Aversion and Scale Compatibility in Two-Attribute Trade-Offs. *Journal of Mathematical Psychology*. 2002;46:315–37.

Bryce, C. L., Loewenstein, G., Arnold, R. M., Schooler, J., Wax, R. S., & Angus, D. C. (2004). Quality of Death: Assessing the Importance Placed on End-of-Life Treatment in the Intensive-Care Unit. *Medical Care, 42*(5), 423-431.

Buitinga L, Braakman-Jansen L, Taal E, van de Laar M. Construct validity of the interview Time Trade-Off and computer Time Trade-Off in patients with rheumatoid arthritis: A cross-sectional observational pilot study. *BMC Musculoskelet Disord*. 2012;13(1):1-7.

Burström, K., Johannesson, M., & Diderichsen, F. (2001). Health-related quality of life by disease and socio-economic group in the general population in Sweden. *Health Policy, 55*(1), 51-69. doi: 10.1016/s0168-8510(00)00111-1

Carstensen LL, editor. Motivation for social contact across the life span: A theory of socioemotional selectivity. Nebraska Symposium on Motivation; 1993.

Carstensen LL, Pasupathi M, Mayr U, Nesselroade JR. Emotional experience in everyday life across the adult life span. *Journal of Personality and Social Psychology*. 2000;79(4):644.

Carter, W. B., Bobbitt, R. A., Bergner, M., & Gilson, B. S. (1976). Validation of an interval scaling: the sickness impact profile. *Health Services Research, 11*(4), 516.

Chang W, Collins E, Kerrigan C. An Internet-based utility assessment of breast hypertrophy. *Plastic Reconstructive Surgery*. 2001(108):370-7.

Chen Y, Ma X, Pethtel O. Age differences in trade-off decisions: Older adults prefer choice deferral. *Psychology and Aging*. 2011;26(2):269.

Churchill DN, Morgan J, Torrance GW. Quality of life in end-stage renal disease. *Peritoneal Dialysis International*. 1984;4(1):20-3.

Churchill DN, Torrance G, Taylor DW, Barnes CC, Ludwin D, Shimizu A, et al. Measurement of quality of life in end-stage renal disease: the time trade-off approach. Clinical and investigative medicine *Medecine Clinique et Experimentale*. 1987;10(1):14-20.

Devlin, N., Tsuchiya, A., Buckingham, K., & Tilling, C. (2009). *Does the value of quality of life depend on duration?* . Report No. 09/07. Department of Economics, City University London. London, UK.

Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: Results from a general population study. *Health Economics*. 1996;5(2):141-54.

Dolan P, Roberts J. To what extent can we explain time trade-off values from other information about respondents? *Social Science & Medicine* (1982). 2002;54(6):919-29. Epub 2002/05/09.

Dolan P, Stalmeier P. The validity of time trade-off values in calculating QALYs: constant proportional time trade-off versus the proportional heuristic. *Journal of Health Economics*. 2003;22(3):445-58.

Dolan, P. (2011). Thinking about it: thoughts about health and valuing QALYs. *Health Econ, 20*(12), 1407-1416. doi: 10.1002/hec.1679

Essink-Bot M-L, Stuifbergen M, Meerding W-J, Looman C, Bonsel G, Group V. Individual differences in the use of the response scale determine valuations of hypothetical health states: an empirical study. *BioMed Central Health Services Research*. 2007;7(62).

Evans M, Khunti K, Mamdani M, Galbo Jørgensen CB, Gundgaard J, Bøgelund M, et al. Health-related quality of life associated with daytime and nocturnal hypoglycaemic events: a time trade-off survey in five countries. *Health Qual Life Outcomes*. 2013;11(1):90.

Fisman, D.N. (2005) Health related quality of life in genital herpes: a pilot comparison of measures. *Sex Transm Infect*, 81: 267-270. 10.1136/sti.2004.011619.

Fowler, F. J., Cleary, P. D., Massagli, M. P., Weissman, J., & Epstein, A. (1995). The role of reluctance to give up life in the measurement of the values of health states. *Medical Decision Making, 15*(3), 195-200.

Gigerenzer G, Gaissmaier W. Heuristic decision making. *Annual Review of Psychology*. 2011;62:451-82.

Gold MR, Siegel JE, Russell LB. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996.

Gyrd-Hansen D. Cost-benefit analysis of mammography screening in Denmark based on discrete ranking data. *International Journal of Technology Assessment in Health Care*. 2012;16(03):811-21.

Handler RM, Hynes LM, Nease RF, Jr., Preston-Martin S. Effect of locus of control and consideration of future consequences on time tradeoff utilities for current health. *Quality of Life Research*. 1997;6(1).

Hausman, D. M. (2010). Hedonism and welfare economics. *Economics and Philosophy, 26*(3), 321.

Hazen GB. Adding extrinsic goals to the quality-adjusted life year model. *Decision Analysis*. 2007;4(1):3-16.

Heintz E, Krol M, Levin L-Å. The Impact of Patients' Subjective Life Expectancy on Time Tradeoff Valuations. Medical Decision Making. 2013;33(2):261-70.

Hogarth RM. A perspective on cognitive research in accounting. *Accounting Review*. 1991:277-90.

Hsu PC, Federico CA, Krajden M, Yoshida EM, Bremner KE, Anderson FH, et al. (2012) Health utilities and psychometric quality of life in patients with early- and late-stage hepatitis C virus infection. *J Gastroenterol Hepatol*, 27(1):149–57.

Iqbal I, Dasgupta B, Taylor P, Heron L, Pilling C. Elicitation of health state utilities associated with differing durations of morning stiffness in rheumatoid arthritis. *Journal of Medical Economics*. 2012;15(6):1192-200. Epub 2012/07/19.

Isaacowitz DM, Toner K, Goren D, Wilson HR. Looking while unhappy mood-congruent gaze in young adults, positive gaze in older adults. *Psychological Science*. 2008;19(9):848-53.

Jelsma, J., Hansen, K., de Weerdt, W., de Cock, P., & Kind, P. (2003). How do Zimbabweans value health states? *Population Health Metrics*, 1(1), 1-10. doi: 10.1186/1478-7954-1-11

Johnson E, Häubl G, Keinan A. Aspects of endowment: A query theory of value construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2007;33(3):461-74.

Kattan MW, Fearn PA, Miles BJ, editors. Time trade-off utility modified to accommodate degenerative and life-threatening conditions. Proceedings of the AMIA symposium; 2001: American Medical Informatics Association.

Kovalchik S, Camerer CF, Grether DM, Plott CR, Allman JM. Aging and decision making: A comparison between neurologically healthy elderly and young individuals. *Journal of Economic Behavior & Organization*. 2005;58(1):79-94.

Lanteri A, Carabelli A. What is Behavioural Economics Like? 2007.

Lee GM, Salomon JA, LeBaron CW, Lieu TA. Health-state valuations for pertussis: methods for valuing short-term health states. *Health and Quality of Life Outcomes*. 2005;3(1):17.

Li, Y., Baldassi, M., Johnson, E. J., & Weber, E. U. (2013). Complementary Cognitive Capabilities, Economic Decision-Making, and Aging. *Psychology and Aging, 28*(3), 595-613. doi: 10.1037/a0034172

Lin M-R, Yu W-Y, Wang S-C. Examination of assumptions in using time tradeoff and standard gamble utilities in individuals with spinal cord injury. *Archives of Physical Medicine and Rehabilitation*. 2012;93(2):245-52.

Loomes G, McKenzie L. The use of QALYs in health care decision making. *Social Science & Medicine*. 1989;28(4):299-308.

Luce MF. Choosing to avoid: Coping with negatively emotion-laden consumer decisions. *Journal of Consumer Research*. 1998;24(4):409-33.

Lundberg L, Johannesson M, Isacson DG, Borgquist L. The relationship between health-state utilities and the SF-12 in a general population. Medical Decision Making. 1999;19(2):128-

Ma, X. (2010). *The effects of age, decision styles, and processing strategies on choice deferral in* *trade-off decisions*. Bowling Green State University. Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num=bgsu1280784967

Mather M, Carstensen LL. Aging and motivated cognition: The positivity effect in attention and memory. *Trends in cognitive sciences*. 2005;9(10):496-502.

Maor Y, King M, Olmer L, Mozes B. (2001). A comparison of three measures: the time trade-off technique, global health-related quality of life and the SF-36 in dialysis patients. *J Clin Epidemiol*, 54: 565-570. 10.1016/S0895-4356(00)00338-3.

McNeil B, Weichselbaum R, Pauker S. Tradeoffs between quality and quantity of life in laryngeal cancer. *New England Journal of Medicine*. 1981;305:982–7.

Mohide, E. A., Torrance, G. W., Streiner, D. L., Pringle, D. M., & Gilbert, R. (1988). Measuring the wellbeing of family caregivers using the time trade-off technique. *Journal of Clinical Epidemiology, 41*(5), 475-482.

National Institute for Health and Care Excellence (NICE), "Guide to the methods of technology appraisal," 2013, http://publications.nice.org.uk/guide-to-the-methods-of-technology-appraisal-2013-pmg9/the-reference-case.

Oliver A, Wolff J. Are people consistent when trading time for health? *Economics & Human Biology*. 2014;15; 41-46.

Panidi K, Kirchsteiger G. Essays on the Application of Behavioral Economic Concepts to the Analysis of Health Behavior. 2012. Available at http://difusion.ulb.ac.be/vufind/Record/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/209674/TOC

Payne JW, Bettman JR. Walking with the scarecrow: the information processing approach to decision research. Blackwell Handbook of Judgment and Decision Making. 2004:110.

Pliskin JS, Shepard DS, Weinstein MC. Utility Functions for Life Years and Health Status. *Operations Research*. 1980;28(1):206-24.

Pyne JM, Fortney JC, Tripathi S, Feeny D, Ubel P, Brazier J. How Bad Is Depression? Preference Score Estimates from Depressed Patients and the General Population. *Health Services Research*. 2009;44(4):1406-23.

Rosser, R., & Kind, P. (1978). A scale of valuations of states of illness: is there a social consensus? *International journal of epidemiology*, *7*(4), 347-358.

Sackett D, Torrance G. The utility of different health states as perceived by the general public. *Journal of Chronic Diseases*. 1978;31(11):697-704.

Saigal CS, Gornbein J, Nease R, Litwin MS. (2001). Predictors of utilities for health states in early stage prostate cancer. *J Urol*, 166: 942-946. 10.1016/S0022-5347(05)65869-7.

Schwartz A, Hazen G, Leifer A, Heckerling P. Life goals and health decisions: what will people live (or die) for? *Medical Decision Making*. 2008;28(2):209-19.

Sharma, S., Brown, G., Brown, M., Hollands, H., Robins, R., & Shah, G. (2002). Validity of the time trade-off and standard gamble methods of utility assessment in retinal patients. *British Journal of Ophthalmology, 86*(5), 493-496

Sherbourne CD, Sturm R, Wells KB. What outcomes matter to patients? *Journal of General Internal Medicine*. 1999;14(6):357-63.

Shimizu, F., Fujino, K., Ito, Y. M., Fukuda, T., Kawachi, Y., Minowada, S., ... & Ohashi, Y. (2008). Factors associated with variation in utility scores among patients with prostate cancer. *Value in Health*, *11*(7), 1190-1193.

Slovic P, Finucane M, Peters E, MacGregor DG. Rational actors or rational fools: Implications of the affect heuristic for behavioral economics. *The Journal of Socio-Economics*. 2002;31(4):329-42.

Smith, R., J., O., & A., H. (1999). *A review of methodological issues in the conduct of willingness to pay studies in health care II: administration of CV survey*. (Working Paper 85-86).

Souchek J, Byrne MM, Kelly PA, K OM, Richardson M, Pak C, Nelson H, Suarez-Almazor ME. (2005) Valuation of arthritis health states across ethnic groups and between patients and community members. *Med Care*, 43: 921-928. 10.1097/01.mlr.0000173600.53788.13.

Stalmeier PF, Wakker PP, Bezembinder TG. Preference reversals: Violations of unidimensional procedure invariance. *Journal of Experimental Psychology: Human Perception and Performance.* 1997;23(4):1196-205.

Stamuli E. Health outcomes in economic evaluation: who should value health? *British Medical Bulletin.* 2011;97(1):197-210.

Stiggelbout A, Kiebert G, Kievit J, Leer J, Habbema J, De Haes J. The 'utility' of the time trade-off method in cancer patients: feasibility and proportional trade-off. *Journal of Clinical Epidemiology.* 1995;48(10):1207-14.

Tosteson AN, Kneeland TS, Nease RF, Sumner W. Automated Current Health Time-Trade-Off Assessments in Women’s Health. *Value in Health*. 2002;5(2):98-105.

Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Science*. 1981;211(4481):453-8.

Tversky A, Kahneman D. Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics.* 1991;106(4):1039-61.

van Nooten F, Brouwer B. The influence of subjective expectations about length and quality of life on time trade-off answers. *Health Economics*. 2004;13:819-23.

van Nooten F, Koolman X, Brouwer B. The influence of subjective life expectancy on health state valuations using a 10 year TTO. *Health Economics*. 2009;18:549-58.

van Osch SM. The construction of health state utilities. Leiden, The Netherlands: Leiden University; 2007.

Van Wijck EE, Bosch JL, Hunink MG. Time-tradeoff values and standard-gamble utilities assessed during telephone interviews versus face-to-face interviews. *Medical Decision Making*. 1998;18(4):400-5.

Verschuuren, M. (2006). *Quality adjusted life years and time trade off exercises: exploring methodology and validity.* (PhD), University of Utrecht, Utrecht, The Netherlands.

Versteegh M, Attema A, Oppe M, Devlin N, Stolk E. Time to tweak the TTO: results from a comparison of alternative specifications of the TTO. *Eur J Health Econ*. 2013;14(1):43-51.

Waljee AK, Morris AM, Waljee JF, Higgins PDR. Individual health discount rate in patients with ulcerative colitis. *Inflammatory Bowel Diseases*. 2011;17(6):1328-32.

Weyler E-J, Gandjour A. Empirical Validation of Patient versus Population Preferences in Calculating QALYs. *Health Services Research*. 2011;46(5):1562-74.

Winter, L., Lawton, M. P., & Ruckdeschel, K. (2003). Preferences for prolonging life: a Prospect Theory approach. *International Journal of Aging and Human Development*, *56*(2), 155-170.

Wittenberg E, Prosser LA. Ordering errors, objections and invariance in utility survey responses: a framework for understanding who, why and what to do. *Applied Health Economics and Health Policy*. 2011;9(4):225-41. Epub 2011/06/21.

Zarate, V., Kind, P., & Chuang, L. H. (2008). Hispanic valuation of the EQ-5D health states: a social value set for Latin Americans. *Value in Health*, *11*(7), 1170-1177.

End of Chapter 3 Notes

i. In this context, biased implies that the health state values elicited using the TTO have been shown to be systematically different from health state values elicited using different methods (specifically, where loss aversion has been proposed to be less of an influence).

ii. See, for example, Hausman (2010) for a discussion as to how philosophical arguments may align with the use of health state values provided by particular groups.

iii. To clarify, the respondent's subjective life expectancy (SLE) is the age at which they expect they will die (e.g. 80 years). 'Remaining years' subtracts the respondent's current age from their SLE to indicate the number of years from their SLE that they have left (remaining years) (e.g. 80 subtracted by 23 years if the respondent's age is 23).

iv. To calculate the (two-point) moving average, the mean number of respondents indicating two adjacent TTO values (e.g. 0.1-0.2) is calculated for consecutive pairs of TTO values (i.e. first and second, second and third, etc.). For example, if 3 respondents indicate a TTO value of 0.1 and 7 respondents indicate a TTO value of 0.2, the moving average plotted between 0.1 and 0.2 (i.e. at 0.15) will be 5 respondents ([3+7=10]/2).

v. Non-traders were kept in the main analysis and health state values of one were assigned to these responses, as has been the case in other studies (Churchill et al., 1984, 1987; Handler et al., 1997). These respondents were not excluded from the main analysis as it was thought their values were representative of their preferences for the health state and not protest values that is, they were not failing to properly engage with the TTO task (e.g. see Smith et al., 1999).

vi. It should be noted that in Versteegh et al. (2013), anxiety and depression had the largest impact on TTO values at a 10-year time horizon whereas at a 5-year time horizon the pain/discomfort dimension had the largest negative impact (a finding substantiated by Jelsma et al., 2003 and others).

vii. Another person or group must make decisions about health care when an individual is unable to do so on his or her own.

Chapter 4. Trade-offs Between Quantity and Quality of Life – The Influence of Behavioural Economic Phenomena

4.1 Abstract

Eliciting preferences to gauge how individuals value specific interventions has become a widely adopted strategy across policy contexts, including health care. An underlying assumption of decision theory is that preferences are invariant irrespective of the method used to elicit them, termed procedural invariance. Where preferences are being elicited to inform decisions such as resource allocation, for example, variability in measured preferences due to method of elicitation (rather than underlying preference) is problematic. The time trade-off (TTO), a tool used in health state preference elicitation, has been shown to yield different health state values depending on how the trade-off is presented.

This study examines procedural invariance in health state values elicited using standard and reverse TTOs. In contrast to the standard TTO, the less commonly investigated reverse TTO asks respondents to provide the minimum number of years in an inferior health state at which point they would be indifferent with living for a specified time in full health. The standard TTO is susceptible to loss aversion for length of life (i.e. respondents are hesitant to trade-off duration for better health), resulting in higher health state values. Loss aversion for length of life is not expected in the reverse TTO since respondents are forgoing better health for longevity, resulting in lower health state values. This hypothesis, that the reverse TTO generates lower health state values than the standard TTO, was assessed alongside secondary hypotheses relating to the search procedure by which responses are obtained, namely, that scale compatibility will yield higher values in standard and reverse TTOs in matching than in choice. Also, the relative abilities of search procedures to capture health state severity were explored.

The results of the study confirm and extend the findings of a handful of previous studies which have documented violations of procedural invariance between standard and reverse versions of the TTO. At a 3-year time horizon, consistent violations of procedural invariance between standard and reverse TTOs in the direction predicted by loss aversion were found for both the relatively mild and relatively severe health state.

For the relatively mild health state, back pain, there was weak support for the hypothesis that loss aversion would upwardly influence standard TTO values, as standard TTO values were indeed higher than the reverse TTO values in matching at a 10-year time

horizon. Standard TTO values did not significantly exceed reverse TTO values when a bisection search procedure was used. There was no difference between matching and bisection at either the 10-year or 35-year time horizon, in either the standard or reverse TTO thus refuting the scale compatibility hypothesis that matching standard TTOs and reverse TTO values would be higher in matching than in bisection.

For the more severe health state, unwell, it was hypothesized that increased attention to the health state when using a bisection procedure may mean that bisection is more sensitive/responsive to the severity of the health state. The standard TTO using bisection yielded lower values than matching at both 10- and 35-year time horizons, and for the reverse TTO, bisection generated lower values at the 35-year time horizon - consistent in the direction expected due to scale compatibility. When comparing standard and reverse TTO values in matching, standard TTO values were consistently higher than reverse TTO values. This relationship was not observed in bisection where reverse TTO values were higher than standard at both time horizons. Notably, there was evidence of MET in bisection but not in matching.

Finally, in terms of the search procedures’ sensitivity to health state severity, a much larger difference between values for back pain and unwell was observed in bisection than matching. It is proposed that this may be due to a lesser effect of scale compatibility and loss aversion in bisection than in matching. This finding is concerning in terms of the validity of TTO values elicited through matching.

Overall, the findings of this study highlight the need to better understand why inconsistencies arise from different variations of the TTO since, without this insight, it is difficult to know which values might offer the most accurate reflection of preferences. Loss aversion and scale compatibility influence health state values to varying degrees depending on the TTO design and therefore it is important to decide how to go about managing theses influences: either by avoiding them or through other means, while also considering which TTO format is most easily understood and internalized by respondents. In addition, further research into the reverse TTO construct is desirable given its representativeness of many clinical decision scenarios.

4.2 Background

Eliciting preferences to gauge how individuals value specific interventions has become a widely adopted strategy across policy contexts ranging from land use to environmental issues. In order to elicit preferences from members of the general public in potentially actionable terms, policy-makers have often relied on research tools, including focus groups and direct/indirect value measurement (Keeney et al., 1990). The use of these tools is underpinned by the notion that respondents know what they want and thus their preferences are useful in informing policy decisions.

An underlying assumption of decision theory is that preferences are invariant irrespective of the method used to elicit them, termed procedural invariance. Despite this, a sizeable body of research spanning a number of decision contexts has shown that the way in which questions are asked and responses are retrieved affects our preferences (Tversky and Kahneman, 1981; Knetsch and Sinden, 1984; Tversky et al., 1988; Bostic et al., 1990). The influence of how questions are asked is often referred to in terms of framing effects, for example, whether outcomes are presented as gains or losses. The latter factor, how responses are retrieved, is most often conceptualized as being either choice-based or open-ended.

Where preferences are being elicited to inform particular issues, such as resource allocation, for example, variability in preferences due to method of elicitation (rather than underlying preference) is problematic. Among a set of possible alternative elicitation tools that provide inconsistent preferences, a significant challenge lies in deciding what constitutes the most appropriate method. In selecting a preferred method, a number of characteristics are desirable in that it is easy for respondents to understand and complete, that it is valid (i.e. measuring what it should) and that it yields consistent preferences.

In health care, a subset of cost-effectiveness analysis, cost-utility analysis, is informed using quality adjusted life years (QALYs). QALYs are a composite measure of survival duration (length of life) and health-related quality of life (hereon HRQoL, usually quantified on a scale from 0 to 1). Preferences may be elicited from various population groups in order to obtain HRQoL values for different conditions (i.e. health states). Health state valuation is an area in which there has been much discussion as to which tools, and which specifications within a given tool, are most appropriate. Among the most frequently implemented tools for eliciting HRQoL is the time trade-off (TTO). Whether or not the TTO elicits values that are both valid and reliable has important implications if values are to be used to inform resource allocation and more specifically whether or not interventions are deemed cost-effective.

4.21 The Standard TTO

In order to arrive at HRQoL estimates (hereon referred to simply as values), the standard TTO model asks respondents to state the number of years they would be willing to give up from a fixed time period in a deteriorated health state in order to live in full health (technically, they are asked to state the number of years in full health that would render them *indifferent* between the two outcomes). The fixed time period, the time horizon, is determined by the researcher and will vary depending on the health state or the objective of the study. For example, a respondent may be asked how many years they would be willing to forgo to live in full health as opposed to living with migraines for 20 years (with immediate death following the 20 years). Consider the respondent indicates indifference between 12 years in full health and 20 years with migraines, their resulting TTO value is 0.6 (0.6=12/20).

4.22 The Reverse TTO

The TTO, as with other methods of preference elicitation in health state valuation and elsewhere, assumes procedural invariance. Thus, it should follow that asking the respondent the reverse question – a reverse TTO – elicits the same values as the standard version.

In the reverse TTO, the respondent is asked to indicate the number of years they would accept in a deteriorated health state as equivalent to a given number of years in full health. Consider the respondent indicates indifference between having migraines for 20 years and living in full health for 14 years in a standard TTO. If they are then asked how many years with migraines they would consider equal to 14 years in full health, the respondent should indicate 20 years, returning to the initial starting point of the standard TTO.

4.23 Possible Sources of Inconsistency

Values elicited from standard and reverse TTOs are subject to several influences that may undermine procedural invariance. Most notable among these influences is loss aversion. Loss aversion implies that individuals asymmetrically value gains and losses such that if they are asked to *forgo* something in one instance (e.g. sell *x*) and to *acquire* an equal outcome in another (e.g. buy *x*), their valuations (i.e. 'selling' and 'buying' values) would be inconsistent (such that $x_{gain}<x_{loss}$, the buying price would be outweighed by the selling price). Thus, instead of outcomes being evaluated as absolutes, they are interpreted as

gains and losses in relation to a reference point (i.e. the value of *x* is dependent on whether it is being acquired or given up) (Tversky and Kahneman, 1991).

Looking at the standard TTO, if losses loom larger than gains, the respondent will be reluctant to trade-off (lose) life years and therefore their TTO values will be upwardly biased. Suppose a respondent is presented with (Q_1, T_1) (seen graphically in Figure 4-1), which (under TTO model assumptions) constitutes their reference point, the location from which gains and losses are evaluated. The respondent is asked to trade-off a gain in health status from Q_1 to full health (FH) against a loss in life years. If the respondent is not loss averse to forgoing length of life, T_2^1 is the number of years indicated by the respondent, whereby the gain in utility resulting from an increased health status is equal to the loss of utility in terms of life years (Bleichrodt, 2002). If the respondent is loss averse, they will weigh the loss in terms of life years (T_1-T_2^1) more than the gain in HRQoL (Q_1-FH) and will therefore prefer the initial starting point (Q_1, T_1) over (FH, T_2^1). The number of years in full health would need to increase from T_2^1 to T_2^2 in order for the respondent to be indifferent between the two outcomes (Q_1, $T_1 \approx$ FH, T_2^2). As a result, loss aversion will have an upward bias on standard TTO values since (T_2^2/T_1) will exceed (T_2^1/T_1).

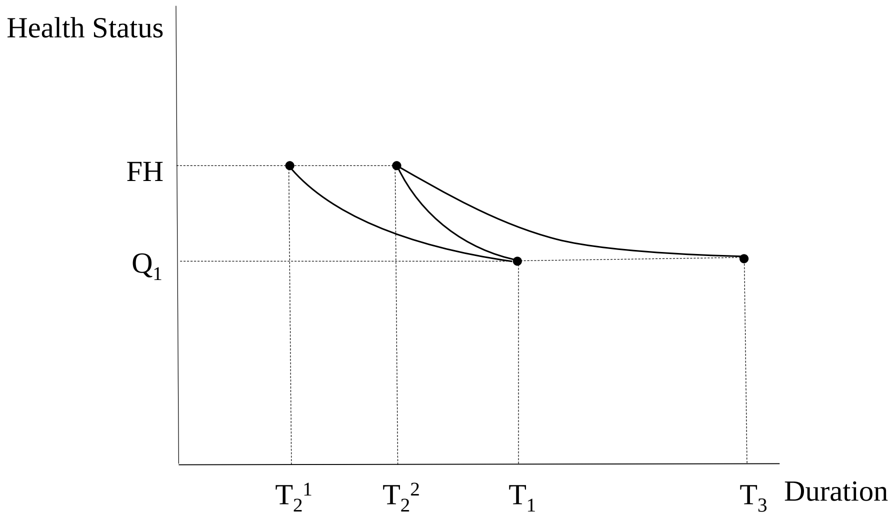


Figure 4-1: Loss aversion in the TTO (adapted from Bleichrodt, 2002) (also Figure 1-5)

Conversely, in the reverse TTO, Bleichrodt (2002) explains that loss aversion may exert a *downward* influence on TTO values. That is, if losses in HRQoL are weighted more than gains in length of life, respondents will demand more years to compensate for the loss in HRQoL. Consider (FH, T_2^2) as the respondent's reference point (i.e. the response provided in the standard TTO). To compensate for the loss in HRQoL, their point of indifference in the reverse TTO will shift from (Q_1, T_1) in the absence of loss aversion to (Q_1, T_3). Since T_2^2/T_3 will be lower than T_2^2/T_1, this means that lower TTO values can be expected in the reverse TTO than in the standard TTO due to loss aversion.

Inconsistencies in the opposite direction to those predicted by loss aversion may be witnessed if there is a maximum endurable time the respondent is willing to live in a severe health state. Maximum endurable time (MET) upwardly influences reverse TTO values and downwardly influences standard TTO values. This is because in the reverse TTO the respondent will be willing to live a limited number of years in the health state before it becomes intolerable, thus decreasing T_1 in the calculation of the TTO value T_2/T_1. Consider, for instance, an individual who is indifferent between 5 years in full health and 10 years with constant migraines in a standard TTO (TTO value is therefore 0.50). In the corresponding reverse TTO, the respondent indicates indifference between 5 years in full health and 7 years with migraines (TTO value of 0.71). In this instance, the reverse TTO value is greater than the standard TTO value because the respondent is unwilling to live any longer than 7 years with migraines, and would prefer fewer years in full health over any longer period with migraines. In the standard TTO, the respondent will be willing to forgo many years to avoid the poor state of health, decreasing their T_2 value relative to T_1. Moreover, MET directly violates the assumed independence between the health state and duration, since the quality weight for the health state will shift from positive to negative as duration increases, implying *longer is not better* (Stalmeier et al., 2001, 2007). MET will not have an impact on TTO values if the respondent anticipates being able to cope with the health state. In terms of time horizon effects, if the respondent believes that there is a maximum period of time that they can live in the health state (e.g. 3 years), then an increase in the time horizon will result in lower TTO values since 3 years in *health state x*/10 years in full health (0.3) is greater than, for example, 3 years in *health state x*/30 years in full health (0.1).

In the standard TTO, loss aversion may be strengthened by the effect of scale compatibility when values are elicited through a matching procedure (whereby the respondent indicates their indifference value without any prompting from the researcher). Scale compatibility is defined as the "dimensional compatibility between input and output" (Selart, 1996, p. 114). In matching, since the response scale is the number of years, it is this attribute (i.e. length of life) that is expected to receive greater weight. Therefore, with matching in the standard TTO, the focus on life years due to scale compatibility will work alongside the effects of loss aversion to increase standard TTO values. In the reverse TTO, scale compatibility suggests fewer years are needed to compensate for the loss in HRQoL, also exerting an upward influence on TTO values (Bleichrodt and Pinto, 2002; Attema and Brouwer, 2013). Attema and Brouwer (2013) explain, "[t]his is because the number of years matters but not the quality, and, hence, respondents make sure that the number of years is close in each case, disregarding their differences in quality." Further, since the response scale is consistent between standard and reverse TTOs, scale compatibility is not expected to influence either version more than the other in matching tasks.

Table 4-1: Directional influences of scale compatibility, loss aversion, and maximum endurable time (MET)

		Standard TTO	**Reverse TTO**
Matching	Scale Compatibility	Up	Up
	Loss Aversion	Up	Down
	MET	Down	Up
Choice	Scale Compatibility	Relatively less than matching	Relatively less than matching
	Loss Aversion	Up	Down
	MET	Down	Up

If instead standard TTO values are elicited through *choice*, the respondent selects among alternatives presented to them (i.e. there is no response scale) and thus less of an impact of scale compatibility than in matching is expected (reasons for this are elaborated on below). Table 4-1 shows the directional influences for scale compatibility and loss aversion across search procedures and TTO constructs.

Finally, the TTO may be influenced by variations in time preference (i.e. how *when* an outcome happens or a good is consumed affects its value). The TTO model is based on the strong assumption of linearity over time (Pliskin et al., 1980). This means that the value of the health state is independent from the time horizon from which it is elicited. A number of studies eliciting time preference for health states, however, have reported that respondents attach a lower value to future outcomes, meaning they are traded-off with greater ease, resulting in lower TTO values (e.g. Redelmeier and Heller, 1993; Cairns and van der Pol, 2000; Attema and Brouwer, 2008, 2012, 2013). The opposing trend, increasing TTO values at longer time horizons, has also been reported and may occur if the respondent anchors on the length of life dimension, due, for example, to scale compatibility, and fails to sufficiently adjust for losses in quality of life (Tversky and Kahneman, 1974; Bleichrodt, 2002).

It is generally accepted that the TTO model assumes linear utility for duration. If TTO values are not in fact independent of the time horizons from which they are elicited, non-linearity may be accounted for to some extent by adjusting values for discounting. Attema and Brouwer (2008, p. 880) explain that although discounting will not cause standard and reverse TTO values to differ, “it can influence the magnitude of any difference that results from other biases. To what extent the magnitude of this difference is caused by discounting can only be assessed by correcting for it.” Given that respondents may differ in terms of the degree and direction (e.g. some respondents may prefer worse outcomes sooner, and other respondents, later) of their discounting, applying a standard discount rate across all respondents may not be appropriate.

As the number of studies that have set out to compare standard and reverse TTO values is relatively small, this study had an overall aim of substantiating the existing evidence and gaining insights into observed inconsistencies. Testing whether procedural invariance holds between values elicited using the standard and reverse TTO is an important aim given the implications of inconsistent values in terms of the TTO’s validity and usefulness as a tool to generate health state values. The reverse TTO, although far less researched, also presents a potentially more intuitive trade-off scenario. For example, the reverse TTO is illustrative of a commonly discussed clinical scenario (e.g. in oncology) that involves a trade between proximal death and prolonged life at the cost of quality of life. So, from this perspective, a better understanding of the nuances associated with this version of the TTO should be explored.

The remainder of this chapter is structured as follows: a review of the relevant literature, presentation of a standard vs. reverse TTO analysis for a mild health state, presentation of a standard vs. reverse TTO analysis for a relatively more severe health state, a sensitivity test for the search procedures comparing data from the mild health state analysis with data from the severe health state analysis, and finally a discussion of the findings and broader implications.

4.24 Other Studies Comparing Standard and Reverse TTOs

A limited number of studies employing varied search procedures and time horizons have sought to assess procedural invariance in relation to the values elicited using standard and reverse TTOs. Table 2 contains an overview of these studies and their findings. Spencer

(2003) observed mixed results in terms of consistency between values, reporting significantly different standard and reverse TTO values in the direction predicted by loss aversion for one of two health states. The smaller (non-significant) differential was found in the more severe health state. Spencer speculated that some respondents may have had MET in the reverse TTO, upwardly influencing their reverse TTO values such that they that were comparable to those elicited in the standard TTO (hence the lack of significant difference).

Bleichrodt et al. (2003) found significant differences between standard and reverse TTO values in 3 of 5 TTOs, specifically those with shorter time horizons (13, 19, and 24 years) and not at the longer time horizons (31 and 38 years). In a second experiment, they observed significant differences between standard and reverse TTO values in 4 of 6 TTOs (two health states each evaluated at three time horizons: 13, 24 and 38 years). As in the first experiment, differences were observed at the two shorter time horizons for each health state but not the longest. As one possible explanation for this finding, a number of other studies have shown that loss aversion in terms of forgoing life years may decrease at longer time horizons, and thus the respondent is then more willing to substitute longevity for improved HRQoL (Bleichrodt and Pinto, 2002; Attema and Brouwer, 2013).

Table 4-2: Investigations comparing standard and reverse TTOs

Study & Author	Subjects	Health State*	Time Horizon(s)**/ Search Procedure	Results
Spencer (2003)	N=40	22232 21222	Standard: 10 Reverse: 2	Mixed. Significant differences in one of two health states. Evidence of LA, SC, MET
			Choice	
Bleichrodt et al. (2003)	N= 65	(expt.1) back pain (expt.2) 22122 22322	Standard: (expt.1) 13, 19, 24, 31, 38 (expt.2) 13, 24, 38 Reverse: reflexive	(expt.1) Significant differences at 13-, 19- and 24-year (expt.2) Significant differences at 13- and 24-year durations LA at shorter durations
			Choice	
Attema and Brouwer (2008)	N=56	Back Pain	Standard: 14, 27 Reverse: 10, 22	Significant differences at all durations Evidence of LA, no MET (mild HS)
			Matching	
Attema and Brouwer (2012)	N=76	Back Pain	Standard: 3, 10, 15, 31, 46 Reverse: reflexive	Significant differences at 3 years LA only at 3 years, no MET (mild HS)
			Choice	
Attema and Brouwer (2013)	N=51	Back Pain	Standard: 3, 10, 31 Reverse: reflexive	Matching: significant differences at all durations. Choice: significant differences at 3 years LA in all matching and 3 years in choice, no MET (mild HS)
			Matching and Choice	
Oliver and Wolff (2014)	N=30	Migraine (varying severity)	Standard: 20 Reverse: reflexive	Significant differences at all levels of severity Evidence of LA, no MET
			Matching	

LA: Loss aversion; SC: Scale compatibility; MET: Maximum endurable time; HS: Health state

*Note: Spencer (2003) and Bleichrodt et al. (2003) described health states using the EQ5D descriptive system, outlined in the methods section of this chapter. Bleichrodt et al. (2003) had respondents complete a table with their responses, allowing them to review and change responses to the different time horizons.

**The term 'reflexive' implies that the number of years in good health indicated by the respondent in the standard TTO was subsequently used in the reverse TTO. That is, if a respondent indicated indifference at 5 years in full health in an 8-year standard TTO, a value of 5 was inputted as the given number of years in full health in the reverse TTO. In Spencer (2003) and Attema and Brouwer (2008), reverse TTO values were elicited instead from fixed reference points instead of based on standard TTO values. This is perhaps a weaker test of procedural invariance than the reflexive version given that values may be affected by discounting (which cannot be assumed to be symmetrical between the two procedures).[i]

In two tests, Attema and Brouwer (2008) found violations of procedural invariance in the direction predicted by loss aversion. As in Bleichrodt et al. (2003) the difference between standard and reverse TTO values decreased as the time horizon increased. In a subsequent study, Attema and Brouwer (2012) adopted a different methodology and reported significantly higher standard than reverse TTO values at only the shortest (3-year) of five time horizons. In both Attema and Brouwer (2008) and (2012), significant differences between standard and reverse TTO values persisted once values had been adjusted for discounting. Recently, Oliver and Wolff (2014) evaluated various degrees of migraine severity (two migraines per week, per month and per three-month period) using a 20-year TTO. They found violations of procedural invariance also in the direction predicted by loss aversion in each scenario.

It has been suggested that one possible determinant in whether procedural invariance is observed is the search procedure (Attema and Brouwer, 2012). Interestingly, although recognizing that their methodologies differ on a number of dimensions, the studies described above that implemented a matching procedure (Attema and Brouwer 2008; Oliver and Wolff, 2014) have consistently found significant differences between standard and reverse TTOs, whereas studies that used choice-based procedures (Spencer, 2003; Bleichrodt et al., 2003; Attema and Brouwer, 2012) have shown mixed results.[ii] Only one study has assessed the relationship between standard and reverse TTO values using different search procedures. In their most recent investigation, and that most pertinent to the current study, Attema and Brouwer (2013) found standard and reverse TTO values differed significantly when a matching technique was used while differences in values elicited using a choice technique reached significance at only the shortest duration (3 years).

In particular, this study served to extend the work of Attema and Brouwer (2013) by evaluating the influence of the search procedure on standard and reverse TTO values. If the search procedure does in fact mediate whether consistent values are observed then this may give reason for preferring one method over the other. This study is novel in evaluating procedural invariance and the effects of the search procedure using health states of differing severity. This specific test of procedural invariance is focused on assessing the validity of TTO values generated using different procedures – i.e. Are different versions of the TTO equally able to measure the construct it is intended to? Are the different constructions equally successful in distinguishing between health states of different severity?

4.3 Hypotheses

Adopting an analytical framework similar to that of Attema and Brouwer (2013), and based on the existing literature of standard and reverse TTOs as well as what is known about the influences that may undermine procedural invariance, three propositions were made:

The scale compatibility hypothesis: Scale compatibility is anticipated to have a relatively small impact on choice valuations compared to matching – whereas scale compatibility is expected to exert an upward influence on matching values in both standard and reverse TTOs. Thus, higher values can thus be expected in matching relative to choice.

Due to scale compatibility, standard and reverse TTO values elicited through matching will exceed standard and reverse TTO values elicited through choice.

The loss aversion hypothesis: Earlier studies have shown that for both choice and matching, reverse TTO values are lower than standard TTO values due to the opposing effects of loss aversion in each construct. This study aimed to substantiate these findings across both a mild and a relatively more severe health state and two different search procedures.

Standard TTO values are higher than reverse TTO values due to loss aversion.

The internal consistency hypothesis: There is evidence that suggests when using monetary outcomes, choice elicits values that are more internally consistent (i.e. standard and reverse values differ to a smaller degree) than matching. More specifically, greater consistency and fewer errors (less variation over repetitive tasks) have been observed in preferences elicited through choice (e.g. see Bostic et al., 1990; Schmidt and Hey, 2004; Attema and Brouwer, 2013).

There is no a priori reason to believe that matching values will differ more than choice values. While scale compatibility is expected to influence matching values and not choice, it should not affect the consistency of standard and reverse TTO values in matching due to its upward influence in both TTO constructs. This suggests that other important differences exist between elicitation procedures that contribute to superior internal consistency in choice values. Attema and Brouwer (2013) are the only investigators to date who have shown that this observation carries over into a health context using standard and reverse TTOs. Thus, refining the loss aversion hypothesis, a further objective was to replicate Attema and Brouwer's observation in a mild health state as well as in a more severe health state.

The internal consistency between standard and reverse TTO values is greater in choice than in matching.

4.4 The Study

The study is presented in detail in 3 parts. The first part of the study (Part A) examined standard and reverse TTO values for back pain, a health state that has been frequently valued in the studies in Table 1. TTO values were elicited through both choice and matching search procedures. The results of a 3-year TTO are presented briefly prior to reviewing the results from the other time horizons.

Part A served primarily to build on the findings of Attema and Brouwer (2013) as they are the only authors to have elicited both choice and matching values in the same study.

Part B had the same objectives as Part A, except that a relatively more severe health state was valued. Only three studies of standard and reverse TTOs have included health states of varying severity in their methodology: Spencer (2003), Bleichrodt et al. (2003), and Oliver and Wolff (2014), although they did not look at the effects of the search procedure on value consistency. This is the first study to investigate the effect of search procedure on standard and reverse TTO values using a relatively more severe health state (e.g. compared to back pain).

The third part presents an analysis carried out using data from Parts A and B. The aim of this analysis was to assess the respective search procedures sensitivity to changes in health state severity and is explained in greater detail below.

4.5 Methods

Face-to-face interviews were carried out, with responses recorded via Qualtrics Survey Software. Each interview lasted approximately 30 minutes. One hundred respondents were each assigned to one of four groups (n=25), described below.

A between-subjects, two-by-two design was used. Respondents were allocated to one of four groups such that a different respondent group completed TTOs using either a matching or a bisection (choice) technique (described below) for either the back pain (Part A) or unwell health state (Part B) (see Figure 4-2). A between-subjects design was used since it was thought that such a design would minimize participant fatigue. Matching and bisection were chosen as the search procedures since they have been the most commonly used in other research of standard and reverse TTOs. Further, the potential for anchoring biases to affect upward and downward titration procedures (e.g. see Lenert et al., 1988) offers justification for the use of bisection as the choice-based procedure.

Part A *Back Pain*	
1st group	Matching 3-, 10-, and 35-year standard and reverse TTO elicitations
2nd group	Bisection 3-, 10-, and 35-year standard and reverse TTO elicitations
Part B *Unwell*	
3rd group	Matching 3-, 10-, and 35-year standard and reverse TTO elicitations
4th group	Bisection 3-, 10-, and 35-year standard and reverse TTO elicitations

Figure 4-2: The elicitation tasks

Both the back pain (Part A) and unwell (Part B) health states were described based on the EuroQol (EQ5D) framework. The EQ5D is a generic preference-based measure of health that describes health-related quality of life over 5 dimensions: mobility, self-care, usual activities, pain and anxiety using different levels of functioning (no problems, slight problems, moderate problems, severe problems and unable to/extreme problems).

Box 4-1: Description of the back pain health state

Frequent back pain means you will:

- have some problems in walking about
- have no problems washing or dressing yourself
- have no problems with your usual activities
- have no pain or other discomfort
- not be anxious or depressed

Back pain was selected on the basis that it is the most frequently investigated health state in the studies of procedural invariance in Table 4-1 and given that is relatively easy for respondents to envisage implications of back pain on a day-to-day basis in their lives. Using a formal EQ5D description to describe back pain (see Box 4-1) enabled direct comparability with other studies (e.g. Attema and Brouwer, 2013, from whom this description is drawn). When asked to value the more severe health state (Box 4-2), respondents may have more difficulty in anticipating the potential implications on their own lives given that they are likely to have less personal experience in living with serious health problems. The rationale behind using an unlabelled health state is as follows. The challenge in assessing a more severe health state is to convey its severity to respondents where most or none of whom will ever had experience of with it. To this extent, it is also important that respondents' valuations are not clouded by others' unique experiences with the health state should they recognize the label used to describe the state (e.g. cancer).

Box 4-2: Description of the unwell health state

Feeling unwell means you will:

- spend much of your time at home, for the most part unable to get out to accomplish daily tasks or attend social engagements
- have slight problems with self-care (i.e. you are unable to complete these activities on occasion because of pain and/or tiredness)
- have a low mood (i.e. lack of ambition to do even basic activities) as a result

Each respondent completed three standard and three reverse TTO valuations. Respondents were asked to provide the *very* least number of years at which point they would prefer the full health outcome (in the standard TTO) or the outcome in the worse health state (in the reverse TTO).[iii] The time horizons for the standard elicitations were 3, 10, and 35 years. A 10-year time horizon was included since it is a frequently implemented and standard time horizon in the derivation of value sets in a number of countries that use the EQ5D protocol (Arensen and Trommald, 2005; Attema and Brouwer, 2012, 2013; Boye et al., 2014). It was thought that whether procedural invariance between standard and reverse TTO is violated at the longer, 35-year time horizon may depend on which search procedure was used. While some studies using a choice procedure to elicit values (e.g. Bleichrodt et al., 2003; Attema and Brouwer, 2008, 2013) have shown decreased standard-reverse value differentials at longer durations, Attema and Brouwer (2013) found significant differences at a 31-year time horizon using a matching procedure. Importantly, it was anticipated that all of the respondents could expect to live at least 35 more years and thus this constituted a reasonable duration over which trades could realistically be made.

Respondents in the bisection condition were taken through an iteration procedure in the standard manner. That is, if the respondent said *no* to a particular choice set, this would decrease the difference in life years between the two choices in the subsequent choice set whereas *yes* to a particular choice set would increase this difference (e.g. suppose the respondent said *no* to accepting 6 years in full health to 10 years with back pain, 6 years would increase to, say, 8 years). Figure 4-3 shows a standard TTO 10-year bisection iteration (see Appendix 4-1 for the decision trees used in the 3- and 35-year standard TTO bisection questions).

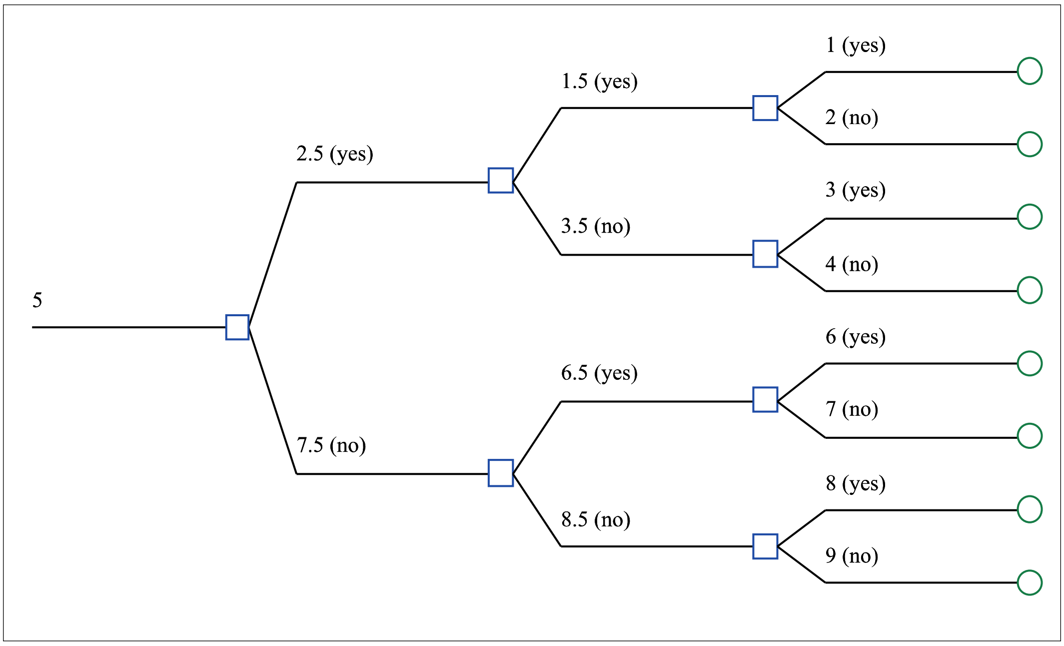


Figure 4-3: A 10-year standard bisection iteration
Respondents were initially offered 5 years in full health or 10 years either unwell or with back pain. Supposing the respondent rejects the 5 years (hence prefers a longer duration in

worse health), they are subsequently offered 7.5 years in full health or 10 years in worse health. If instead the respondent accepts the initial offer of 5 years in full health, their subsequent offer would be 2.5 years in full health or 10 years in worse health.

For the reverse TTO bisection questions, the initial time horizon was calculated as one-half of the interval between the respondent's age (suppose, 20 years) and the age to which the respondent anticipates living (their subjective life expectancy - SLE, suppose, 80 years). In this instance, the respondent would be offered a choice between [FH, 20] and [back pain, 30], where 30 years is calculated as [80-20]/2, i.e. one-half of the 60-year interval between respondent age and SLE, and 20 is number of years in full health provided by the respondent in the standard TTO. If the respondent opts for the full health outcome in preference to 30 years with back pain, the number of years with back pain would be increased from 30 to 45, i.e. the midpoint of the 30-year interval between the 30 years in back pain that was declined and the SLE. On the other hand, if the respondent opts for the back pain outcome over full health for 20 years, the number of years with back pain is decreased from 30 to 15, i.e. the midpoint of the 30-year interval between the 30 years in back pain that was declined and the respondent's current age (15=30-[60-30]/2). Succeeding iterations follow the same bisection pattern based on the respondent's choices. Note that all respondents had bisection trees tailored depending on their SLE and standard TTO responses.

The respondents who completed TTO valuations using a matching procedure stated their points of indifference in standard and reverse TTOs that were then verbally confirmed by the interviewer. Figure 4-4 shows the 10-year TTO question for back pain where values are elicited through matching.[iv]

You have been to the doctor recently and received some bad news. You are faced with a shortened life expectancy as well as back pain.

- You have some problems in walking about.
- You have no problems to wash or dress yourself.
- You have no problems with your usual activities.
- You have no pain or other discomfort.
- You are not anxious or depressed

You have two options:

- Have frequent back pain for the remaining 10 years of your life
- Live in full health (i.e. no major health problems) for less than 10 years

Please indicate the very least number of years in good health (less than 10) you would be willing to accept as an alternative to the 10 remaining years of your life with frequent back pain.

I consider living **______** years (less than 10) in full health better than living the remaining **10 years** of my life with frequent back pain.

Figure 4-4: A standard-frame, matching TTO question for back pain

This finding adds evidence against procedural invariance at very short durations, a result that previous studies have attributed to greater loss aversion at shorter time horizons in the standard TTO where preferences may be lexicographic (i.e. respondents are unwilling to forgo longevity for improvements in health status when length of life is short) (McNeil et al., 1981; Miyamoto and Eraker, 1988; Pliskin et al., 1980; Bleichrodt et al., 2003; Attema and Brouwer, 2012, 2013).

All of the respondents completed a 25-year TTO practice question using a search procedure congruent with their main elicitations (e.g. those in the bisection group completed the 25-year TTO using a bisection procedure) to ensure that they understood the TTO task. The starting point in the reverse TTO was the number of years in full health the respondent had provided in the standard TTO question (thus the assumption that values are reflexive could be made). The main elicitation questions were interspersed with other non-valuation questions in order to decrease the chance of respondents recalling the durations used in the standard trade-off values in the reverse TTO elicitations.

In instances where the respondent indicated *yes* or *no* to all of the offers for a given TTO valuation, whether they would be 'always willing to accept' or 'never willing to accept' was clarified by the interviewer (the constraints on the bisection procedure are such that without asking the respondent directly, these values cannot be inferred). Therefore, in both choice and matching, values of 0 and 1 were possible. In the standard TTO, if a respondent is always willing to trade, this may imply that the health state has a value of 0 and thus the respondent perceives it to be worse than death. In this case, the TTO in its conventional form is not appropriate and other tools should be used in its place (Torrance, 1986). On the other hand, a respondent who is 'never willing to trade' is implying that the health state has a value of 1. In the reverse TTO a value of 0 was assigned to a health state the respondent was 'never willing to accept' trading into. Conversely, a value of 1 was assigned to a health state for which the respondent was not willing to forgo time ('never willing to accept' in the standard TTO) and to a health state the respondent was 'always willing to accept' in the reverse TTO.

4.51 Discounting

Individual discount rates were calculated implementing a "delay of ill health" method based on that used by Cairns (1992) and van der Pol and Cairns (2000), paralleling the approach used in the study in Chapter 2 (see Appendix 2-2). Attema and Brouwer (2008, 2012, 2013) have been the only authors to assess the effects of discounting in the context of standard and reverse trade-offs. They found that the majority of respondents had positive discount rates, thus adjusted values were higher than unadjusted values.

4.52 The Samples

One hundred respondents were recruited through the Behavioural Research Laboratory (BRL) at the London School of Economics. Respondents received remuneration of £5 for their participation. The respondent pool available from the BRL encompasses largely higher education students of various nationalities although it is accessible without restrictions to anyone over the age of 18 (e.g. LSE students, students from other universities, LSE staff members, individuals working in the local area or who are regular lab participants in paid studies across London). The sample is summarized in Table 4-3.

Table 4-3: Respondent characteristics

	Back Pain Health State		Unwell Health State	
	Matching	*Bisection*	*Matching*	*Bisection*
N	25	25	25	25
Sex				
Male	5	16	19	8
Female	20	9	6	17
Age (mean)	25 (19-39)	24 (19-35)	23 (18-30)	24 (20-36)
Education				
Secondary School	1	1	0	1
Undergraduate	11	16	9	12
Advanced Degree	13	8	16	12

Results of the discounting exercise: The median discount rates are presented in Table 4-4. The majority of respondents in each of the four groups had positive discount rates, implying that they place relatively less value on future years (i.e. in delaying the beginning of their period of ill health they would be willing to accept a higher number of days spent in the poorer state).

Table 4-4: Median discount rates (%) and directionality (number of respondents)

	Median Discount	Positive	Negative	Equal
Unwell				
Matching	1.0% (-3 to 4)	14	2	9
Bisection	1.0% (-19 to 5)	15	5	10
Back Pain				
Matching	1.0% (-3 to 5)	17	2	6
Bisection	1.1% (-8 to 17)	17	5	3

The observed discount rates are comparable to those found by Dolan and Gudex (1995; –2.9 to 1.4%), van der Pol and Roux (2005; 0.0 to 2%) and Cairns (1992) (-0.1 to 0.5%). West et al. (2003) also found that among samples of the general public and health professionals, the majority of the discount rates calculated from 4 different scenarios had medians of zero. Cairns (2000) remarked that the period of delay that is provided in the discounting task may affect discount rates; specifically, lower rates may be observed at longer periods of delay (Redelmeier and Heller, 1993; West et al. 2003). Other possible explanations for low discount rates are returned to in the limitation section.

The analyses presented below are of *adjusted* TTO values. Given that the majority of discount rates did not differ significantly from zero, it is worth noting no marked differences when the same tests were undertaken using *unadjusted* TTO values – i.e. where significance was reached using the adjusted values it was also observed when analyses were carried out using the unadjusted values. The unadjusted TTO values and results are in Appendix 4-2.

4.53 Analyses

Quantitative analyses were carried out in SPSS v. 22. Shapiro-Wilk tests were used to test for the normality of each TTO elicitation (Shapiro and Wilk, 1965). Shapiro-Wilk tests assess the squared Pearson's correlation coefficient between the respondents' values and those of a normalized distribution and are particularly powerful when using small samples (i.e. approximately 3-50 respondents). The tests indicated that the majority of the data were non-normally distributed (12 of 16 tests showed non-normal data distributions, $p<0.05$); thus, the majority of the analyses focused on median values. Moreover, lending further support to the use of median values, mean values of small samples may be skewed by outliers. Differences were considered statistically significant at $p<0.05$.

Mann-Whitney U tests (which assess the null hypothesis that each group in the study has the same median) were conducted to assess whether standard and reverse TTO values were different depending on the search procedure (e.g. whether the 10-year standard TTO value for matching differed from the 10-year standard TTO value for bisection). Wilcoxon signed ranks tests were conducted to determine whether standard TTO values differed significantly from reverse TTO values.

4.6 Results: 3-year TTO (Back Pain and Unwell)

To simplify the presentation of results, the findings from the 3-year elicitations are presented in Appendix 4-2. In brief, the 3-year TTO elicitations showed violations of procedural invariance in the direction predicted by loss aversion that were significant in both health states and when using either search procedure.

4.7 Results: Part A (Back Pain)

The intention of Part A was to substantiate the results of Attema and Brouwer (2013) who are the only authors to have concurrently assessed choice and matching search procedures in the context of standard and reverse TTOs. Fifty respondents were asked to complete a series of TTOs for back pain, described in Box 4-1. Median values for the back pain TTOs are seen in Table 4-5.

Table 4-5: Median TTO values for back pain

	Standard		Reverse	
	10-year	35-year	10-Reverse	35-Reverse
Matching				
Median	0.82	0.86	0.67	0.79
IQR	0.64-0.91	0.57-0.96	0.48-0.78	0.56-0.95
Bisection				
Median	0.73	0.86	0.65	0.80
IQR	0.54-0.96	0.52-0.95	0.48-0.89	0.58-0.92

IQR = inter-quartile range.
Note the following missing values due to computer errors in recording responses (namely in the computation of individual bisection trees): Back Pain Bisection, n=24 for 10-year reverse; Back Pain Matching, n=24 for 35-year reverse. Non-traders: Back Pain Bisection, 3 non-traders for 10-year reverse, 2 non-traders for 35-year reverse. When non-traders are removed from the analysis, the significance results remain unchanged.

4.71 Scale Compatibility Hypothesis

The effects of scale compatibility were examined by looking at how standard choice compared with standard matching values and how reverse choice compared with matching values, respectively. Scale compatibility should upwardly influence matching values so that they are higher than choice (bisection) values; however, matching and bisection values did not differ in either standard or reverse TTOs or at either time horizon.

4.72 Loss Aversion Hypothesis

Standard TTO values were expected to be higher than reverse TTO values due to loss aversion. Table 4-6 shows the significance values between standard and reverse values as indicated by the Wilcoxon tests.

Table 4-6: Wilcoxon Signed Rank Tests, p values for differences between standard and reverse TTO values for back pain

	TTO 10	TTO 35
Matching	0.007	ns (0.86)
Bisection	ns (0.07)	ns (0.20)

Note: ‘ns’ indicates the difference was not significant

At face value, all matching and bisection values are higher in the standard TTO than in the reverse TTO, but the only significant difference to emerge was between the standard and reverse values in the 10-year TTO values elicited through matching (p=0.007). Given that there was a trend towards the 10-year TTO reaching significance in the bisection group, only weak support can be offered towards greater procedural invariance in choice.

4.73 The Internal Consistency Hypothesis

The above findings add to those of Attema and Brouwer (2013) in that procedural invariance held in bisection whereas matching TTO values were procedurally invariant at only the longer, 35-year time horizon. Given that this support is based on a single time horizon, it would be beneficial for further research to carry out elicitations using the different search procedures at intermediate time horizons.

4.8 Results: Part B (Unwell)

Part B was conducted with the purpose of examining trade-offs across search procedures using a relatively more severe health state. A different group of 50 respondents than in Part A completed TTO elicitations in Part B.

The three studies (Spencer, 2003; Bleichrodt et al., 2003; Oliver and Wolff, 2014) that have evaluated somewhat more severe health states have shown mixed results in terms of procedural invariance. In evaluating a more severe health state, MET becomes a relevant consideration that may impact on whether procedural invariance is observed. In particular, in the reverse TTO, where the task is concerned with prolonging life in poor health, MET may become a factor more often. Increased attention to HRQoL in choice may mean that this search procedure is more sensitive to severity and thus MET preferences, should they occur.

Table 4-7 shows the median TTO values for unwell. There were no instances in either matching or bisection groups where respondents perceived the health state to be worse than death (thus a TTO value of 0 or less) in the standard TTO.

Table 4-7: Median TTO values for unwell

		Standard		Reverse	
		10-year	35-year	10-Reverse	35-Reverse
Matching					
	median	0.81	0.80	0.60	0.77
	IQR	0.64-0.86	0.67-0.92	0.46-0.68	0.49-0.80
Bisection					
	median	0.43	0.36	0.47	0.50
	IQR	0.30-0.74	0.18-0.66	0.28-0.68	0.00-0.76

IQR = inter-quartile range.
Note: n=23 for 10-year reverse TTO. Unwell Matching: 1 non-trader for 10- and 35-year reverse (same respondent), Unwell Bisection: 5 non-traders for 10-year reverse, 7 non-traders for 35-year reverse. The two minimum values for the bisection group (0.1 and 0) were adjusted for discounting from slightly higher values and thus these respondents were not implying that the state was worse than death.

4.81 Scale Compatibility Hypothesis

As in Part A, Mann-Whitney U tests were carried out to evaluate whether the differences between TTO values elicited from the respective search procedures were significant. As would be expected due to the upward influence of scale compatibility, the tests indicated that bisection values were significantly lower than matching values in the 10-year standard TTO (p=0.001) and the 35-year standard TTO (p=0.001). In addition, the 35-year reverse TTO values for matching were significantly higher than bisection values (p=0.030).

4.82 The Loss Aversion Hypothesis

Looking at Table 4-7, the matching values are in the directed predicted by loss aversion, with standard TTO values higher than reverse values. However, when values were elicited through bisection, standard values are similar or lower than reverse values. Using a matching procedure, differences between standard and reverse TTO values were significant at both time horizons in the direction predicted by loss aversion (10-year, p<0.001; 35-year, p=0.04).

When a bisection procedure was used, significant differences were observed between standard and reverse TTO values only for the 10-year time horizon (both p=0.02), with reverse TTO values exceeding standard TTO values, in contrast to the predicted effects of loss aversion.

The results of the Wilcoxon signed ranks tests evaluating the significance of these observations are show in Table 4-8.

Table 4-8: Wilcoxon Signed Rank Tests, P values for differences between standard and reverse TTO values for unwell

	p-values for Standard-Reverse TTO Differences	
	TTO10	TTO35
Matching	<0.001	0.04
Bisection	0.02	ns (0.93)

Note: ‘ns’ indicates the relationship was not significant

In order to get a better picture of the response patterns, individual data were examined. Spencer (2003) commented that MET has the ability to counteract the effects of loss aversion in the reverse TTO. There was evidence of MET in bisection (5 respondents were unwilling to prolong life in the unwell health state in the 10-year reverse TTO and 7 were unwilling in the 35-year reverse TTO) that was, importantly, not seen in matching (with the exception of a single non-trader in the 10- and 35-year reverse TTOs).[v] This is notable because it shows that matching may be insufficiently able to capture HRQoL.

Table 4-9: Directionality of differences in 10- and 35-year TTO values (v)

	10-year	35-year
Matching		
v(standard) > v(reverse)	21	18
v(standard) < v(reverse)	3	5
v(standard) = v(reverse)	1	2
Bisection		
v(standard) > v(reverse)	14	10
v(standard) < v(reverse)	3	11
v(standard) = v(reverse)	6	4

Note: n=23 for 10-year bisection, N=25 for all others

In Table 4-9, note that in the 35-year TTO for bisection there is a greater number of respondents who valued the reverse TTO *higher* than the standard TTO. This may be due to respondents being reasonably satisfied with 35 years, or adopting a goal close to 35 and misunderstanding that if they asked for any more than 35 years, the entire range (now until 35+*x* years from now) would be spent in the health state.

4.83 The Internal Consistency Hypothesis

Again, tentative support can be offered in terms of greater procedural invariance in bisection than in matching: 1 of the 2 time horizons for bisection (10-years, which notably disappeared when non-trading respondents were excluded from the analysis) and both time horizons for matching violated procedural invariance.

4.84 A 'Sensitivity' Analysis (Back Pain and Unwell)

Based on the proposition that the lack of scale compatibility in bisection focuses respondents to a greater degree on quality of life, an additional exploratory analysis was carried out to test whether bisection or matching is more sensitive to changes in health state severity using the values obtained in Parts A and B. It was anticipated that bisection values would differ more between unwell and back pain than would matching values. This is because a relatively greater strength of scale compatibility and loss aversion was expected in matching (both drawing attention to length of life) and therefore more attention was expected to be paid to the health state in the bisection procedure.

For each search procedure, median unwell TTO values (Table 4-7) were subtracted from those for back pain (Table 4-5) and are shown in Table 4-10. If both search procedures are equally sensitive to HRQoL, then the top and bottom row of the table should show similar values (i.e. the search procedures are equally able to capture the difference in severity between unwell and back pain).

Table 4-10: Differences between median back pain and unwell TTO values by search procedure

		10-year	*35-year*	*10-rev*	*35-rev*
(v)Back Pain- (v)Unwell	Matching	0.01	0.06	0.07	0.02
	Bisection	0.30	0.51	0.19	0.31

The much larger differences between values for bisection suggest that there is greater attention to the HRQoL attribute using this search procedure. Notably, TTO values for both health states were very similar at both durations using a matching procedure: matching effectively generated the same values for back pain as it did for the more severe unwell health state.

One possible explanation as to why matching values were less sensitive to differences in health state severity may be that matching presents a more cognitively demanding task than choice (e.g. that the computation of the respective attributes is particularly taxing) and the respondents in the matching condition may resort to simplified decision strategies. It has been proposed that cognitive easing strategies may be used when trade-offs are difficult to make (Hogarth, 1991; Abhyankar et al., 2013). For example, respondents may base their decision predominantly on the amount of years of life since this dimension is more easily understood and can be emotionally processed more readily than changes in HRQoL (see Hsee's (1996) evaluability principle; Slovic, 2002). Relying solely on one attribute to arrive at their decision is termed lexicographic processing (see Miyamoto and Eraker, 1988).

Further, some studies have suggested that respondents may use a proportional trade-off heuristic in standard TTO tasks (e.g. Dolan and Stalmeier, 2003 and Attema and Brouwer, 2012), although less is known in terms of decision strategies that are used in the reverse TTO. A proportional trade-off heuristic, for example, may imply the respondent requires half (standard) or double (reverse) the number of years they are provided with in the given health state (standard) or full health (reverse) in order to indicate indifference, for example. In looking at the value distributions for matching, this seems like a plausible explanation since responses for many of the standard and reverse TTOs appear to be clustered around a couple of values.

Table 4-11: Review of hypotheses and results

	Hypothesis	Result
The scale compatibility hypothesis	Due to scale compatibility, standard and reverse TTO values elicited through matching will exceed standard and reverse TTO values elicited through choice.	Back Pain: Matching and bisection values did not differ in either standard or reverse TTOs or at either time horizon. Unwell: Matching values were significantly higher than bisection values in the 10-year and the 35-year standard TTOs and the 35-year reverse TTO.
The loss aversion hypothesis	Standard TTO values are higher than reverse TTO values due to loss aversion.	Back Pain: Standard values were higher than reverse values in the 10-year TTO elicited through matching (p=0.007). Unwell: In matching, standard values were higher than reverse values at both time horizons (10-year, p<0.001; 35-year, p=0.04). For bisection, significant differences were observed between standard and reverse TTO values only at the 10-year time horizon (p=0.01), with reverse TTO values exceeding standard values.
The internal consistency hypothesis	The internal consistency between standard and reverse TTO values is greater in choice than in matching.	Back Pain: Matching TTO values differed significantly at the 35-year time horizon but not at 10 years. Unwell: For bisection, a difference was observed in the 10-year TTO (although not when non-trading respondents were excluded from the analysis). For matching, standard and reverse TTO values differed significantly at both 10- and 35-year time horizons.

4.9 Discussion

An assumption of procedural invariance underlies the TTO that, if violated, presents the challenge of understanding which variant yields the most accurate values. Assessing 3 key hypothesis set out in Table 4-11 above, this study provides mixed evidence in terms of procedural invariance with matching showing greater inconsistencies between standard and reverse TTOs than bisection. For the milder health state, back pain, procedural invariance was violated in the 10-year time horizon in matching, while in bisection, standard and reverse TTO values were not significantly different. At the 35-year time horizon no significant differences were witnessed for either bisection and matching. For unwell, significant differences were observed between median values at both the 10-

year and 35-year time horizons for matching, whereas standard and reverse TTO values only differed significantly at the 10-year time horizon for bisection (and did not significantly differ when non-traders were removed). Although Attema and Brouwer (2013) observed that matching values differed significantly at 31 years, it is unsurprising that no significant differences emerged at 35 years in view of the results of other studies that have shown at longer time horizons there is decreased loss aversion and increased attribute interchangeability (McNeil et al., 1981; Bleichrodt and Pinto, 2002; Bleichrodt et al., 2003; Attema and Brouwer, 2008).

Looking at the reverse TTO, if the prospect of being able to live out one's entire life (despite deteriorated health) reorients the respondent to consider, for example, their life expectancy external to the task as their reference point, then reverse TTO values may decrease as the respondent is now loss averse to forgoing years from this external reference point (e.g. 80 years). The notion that respondents may refer to reference points that are not included in the TTO task has been found in a number of studies including those in this thesis (e.g. Chapter 2) and elsewhere (e.g. van Nooten and Brouwer, 2004; van Nooten et al., 2009; Heintz et al., 2013).

It is important to note that it can be difficult to ascertain the role of MET and unwillingness to trade in the reverse TTO and further work, likely requiring qualitative analyses or think-aloud tasks to be conducted to better understand valuations elicited using this method. If MET is strong, respondents will be unwilling to trade increased duration for worse health, eliciting values of 0 in the reverse TTO. Spencer, however, comments that MET may also *upwardly* influence values if the respondent is willing to live *some* amount of time in the health state before deciding it is intolerable. In this instance, MET will have an upward influence on reverse TTO values and as a result, values may appear to be more consistent with or exceed standard TTO values. Consider as an example, two individuals: one individual is indifferent between 8 years in full health and 10 years with constant migraines (but for no period longer than the 10 years, due to MET) therefore has a TTO value of 0.8 (8/10), while a second individual refuses to trade better health for longevity, resulting in a TTO value of 0 (8/0). This is a worthwhile avenue for further research given the reverse TTO's relatively direct applicability to recent advancements across various therapeutic areas (i.e. prolonging life for those with serious conditions but perhaps not for a long period of time).

There is no widely agreed upon strategy for *how* values should be elicited; that is, which search procedure should be used. One possible means of preferring one method over another is greater consistency. Importantly, however, while desirable, internal consistency does not translate into a particular method's superiority in terms of generating accurate preferences. As such Attema and Brouwer (2013) refer to internal consistency as a "necessary, but not sufficient, characteristic of preference elicitation methods." If a particular search procedure yields similar values regardless of the health state, for example, this would seem to compromise its usefulness as a means of measuring health benefit. That is, a tool that elicits consistent preferences would intuitively only seem useful insofar as the preferences (and thus consistency) are not entirely artefacts of the elicitation procedure. In an attempt to provide insight on this issue, an exploratory analysis examined the degree to which bisection and matching values differed across the health states. This study showed very small value differences between the back pain and unwell health states across all time horizons when a matching procedure was used, suggesting that matching is insufficiently sensitive to variations in HRQoL. This study shows that a choice procedure (bisection) appears to be significantly more sensitive to differences in health state severity and thus provides a basis for preferring this procedure for obtaining TTO values over matching.

Several studies have shown that when the standard TTO is asked in the reverse manner, standard values exceed reverse values and are thus not in fact invariant. This observation seems to be mediated to some degree by the search procedure used to elicit values. Although in theory matching and bisection values should not differ, there is limited evidence that standard TTO values tend to be higher than reverse TTO values in choice than in matching. Other authors have noted that underlying differences, beyond those of scale compatibility and loss aversion as they are described above, may exist between search procedures. Bleichrodt et al. (2003) and Sumner and Nease (2001) comment that choice and matching involve different cognitive processes – a notion put forward several decades ago by Tversky et al (1988). Bleichrodt et al. (2003, p.137) remark that "Perhaps, loss aversion is more important in matching than in choice." One possible explanation for this has been that choice-based tasks provide a more neutral decision context than matching. Moreover, in matching, rather than trading off between attributes (e.g. duration and health state), respondents may instead focus on the more prominent attribute (e.g. duration) and anchor and adjust their valuation from this point (Tversky et al., 1988). To this extent, Oliver (2013) suggests that matching tasks might best be avoided in instances where respondents are asked to value unfamiliar outcomes (e.g. health states), since their unfamiliarity may prompt them to rely on simplistic decision strategies (such as relying on a single attribute, duration) to provide their responses.

As the number of studies investigating the reverse TTO is quite limited, several questions remain unresolved with regards to trade-off dynamics using this version of the TTO. In contrast to the standard TTO where there is an imposed limitation of the respondent’s life expectancy, the reverse TTO leaves open the possibility for the respondent to fulfil their initial expectations for longevity if they are willing to accept a decrease in health status. As a result, HRQoL and length of life may be weighted differently than in the standard TTO if, for example, the respondent desires to reach a certain age regardless of the health state. The scenario presented in the reverse TTO may also encourage the respondent to more strongly consider possible life-goals they wish to achieve, which subsequently serve as reference points from which they make a trade-off (Hazen, 2006; Chapter 2 of this thesis). Notably, the latter is inconsistent with the TTO (and QALY) model, whereby trade-offs are expected to be a function of the health state rather than an external objective.

In addition, unlike the standard TTO where the respondent is endowed with two unfamiliar attributes (i.e. an abbreviated life expectancy and a poor state of health), in the reverse TTO it is assumed that they are in their current state of good health that may constitute a stronger endowment effect (whereby the respondent allocates disproportionate value to their status quo, in this case, their current state of health). Qualitative evidence found in Spencer (2003) supports this proposition in that respondents were apprehensive toward trading out of a state of full health in the reverse TTO. Spencer (2003) had respondents imagine that they were to live in a health state inferior to full health for 2 years with the option of living longer in an even lower health state (i.e. HRQoL). Note that this differs from other studies of the reverse TTO which typically involve the respondent indicating their point of indifference between full health for a shorter time horizon (x) or a given, deteriorated, health state for a longer time horizon (x+1) (and both options followed by immediate death). Interestingly, Spencer (2003) found that the effect of endowment was lessened when the respondent was asked to trade into a worse health state from a state other than full health. It may be that when the respondent is endowed with a dimension that is difficult to interpret (i.e. a quality of life with which they are unfamiliar) they are more willing to make trade-offs from this attribute, or, conversely, that people are less willing to reduce something perfect – i.e. full health- than something that is already perceived as damaged (i.e. less than perfect). This may help to explain why there are fewer non-traders in the standard TTOs compared to the reverse TTOs.

4.10 Limitations

There are limitations of the study that should be noted. Firstly, given the reflexive exercise presented by the reverse TTO it might be argued that reverse TTO responses are susceptible to recall effects, whereby the respondent remembers the initial trade-off time horizon and states this as their reverse TTO response. This is unlikely to have occurred since the main elicitation questions were interspersed with filler items place and respondents in the bisection condition did not view their actual response since it was inferred from their choices. Moreover, if recall was more prominent in matching (where respondents may recall their exact point of indifference in the standard TTO) than in bisection, then greater consistency could be expected between standard and reverse TTO values elicited using matching. Additionally, ordering effects, in that standard TTO valuations were always carried out before reverse TTO valuations, and the fact that questions were not randomized between respondents, may have affected respondents' valuations, although the extent or how the ordering of questions might influence values is unclear.

Another important limitation concerns the wording of the trade-off scenarios. Respondents were asked to imagine that they had received *bad news* and that they faced a *shortened* life expectancy – both descriptions which may have induced greater loss aversion than would have a more neutral phrasing. Despite this, given that there were no differences in the questions shown by search procedure or health state, the (potentially upward) impact on values may be assumed to be similar across all respondents.

Finally, a within-subjects design may offer a stronger test of the sensitivity of search procedures to differences in health state severity. This option may be more feasible when fewer time horizons, for example, are being evaluated. In this study, however, evaluation of multiple time horizons was prioritized and the number of questions limited in order to minimize respondent fatigue and maintain respondents' attention.

4.11 Broader Implications

Whether loss aversion or scale compatibility contribute to inconsistencies in TTO valuations is far less contestable than whether these influences should be accepted as reflecting true preferences (and, more generally, incorporated into decision models) or, instead, avoided or corrected (Bleichrodt et al., 2003). Creating decision models which take loss aversion and scale compatibility into account faces several challenges. One challenge is that the influence of loss aversion and scale compatibility appear to vary depending on the construction of the TTO (in terms of its framing as standard or reverse, the search procedure, and the time horizon). This means a significant amount of research would need to be conducted in order to understand the different magnitudes of influences across even a few different constructions.

The latter proposition, to seek to avoid or correct for loss aversion and scale compatibility, suggests that these influences are in fact *biases*. From this perspective, the variation of influence across constructions is favourable, such that it may be possible to identify constructs where the effects of scale compatibility and loss aversion are minimal. It is also worth considering that the effects of other biases (e.g. nonlinearity for time, upward influence) may be offset by loss aversion (downward influence), meaning that values that are uncorrected (i.e. for which no accommodation has been made for the effects of loss aversion) may provide a more accurate representation of underlying preferences than would values where corrections had been put in place (thus unbalancing the two opposing directional influences) (Oliver and Wolff, 2014; Bleichrodt, 2002).

Alternatively, it may be desirable to apply corrective measures; for example, if the use of a particular construct is required or especially relevant to the topic (clinical state) of interest. Oliver and Wolff (2014) suggest two options. The first involves controlling for loss aversion by taking the midpoint of the median standard and reverse TTO values. This proposition rests on the assumption that there is the same degree of loss aversion in both standard and reverse TTOs. Their second suggestion involves obtaining an estimation of the effect of loss aversion in the standard TTO and subsequently applying this estimation as a function of TTO values as follows. Borrowing their notation, if the standard TTO is constructed with a 10-year time horizon and the respondent indicates indifference at 8 years, this would imply that in the absence of loss aversion for length of life (and assuming the loss aversion parameter is roughly 2:1, as has been shown when using monetary outcomes, see Tversky and Kahneman, 1991) that they would be willing to forgo 4 years (instead of 2).

Finally, it is important to consider how respondents answer trade-off questions both for purposes of health state valuation as well as for the increasing number of similar decisions made in real world settings as the population ages and as innovations create greater possibilities for treatments prolonging life at various HRQoL. Some have argued that the reverse TTO presents respondents with a task that is of greater real world plausibility and representativeness of the types of decisions respondents might face in clinical settings (i.e. compared to standard TTO-type questions) (Spencer, 2003; Jansen et al., 1998). That is, an increasing number of decisions are likely to be made with medical advancements alongside aging populations in regards to prolonging (increasing) longevity with potentially decreased levels of HRQoL. These types of questions have already begun to take a more prominent position in debates of health care prioritization and resource allocation and will continue to do so as the number of interventions enabling longer lives, perhaps at the expense of good health, proliferates. Therefore, at both an individual and societal level, the question of whether we wish to live longer if it is not always in better health – i.e. the scenario presented in the reverse TTO – is one that has considerable relevance.

Future research may seek to expand on the implications of asking respondents to trade-off quality for duration, which will involve taking a closer look at the initial endowment of good health. Adopting variations of methods used by Spencer (2003), who had respondents trade from a health state other than full health, for example, may offer a fruitful way forward in terms of assessing the effects of the endowments of varying health states.

The few previous studies of standard and reverse TTOs have been varied in their methodologies and therefore it is difficult to synthesize their findings. This study drew on elements from each of these studies in order to gain a clearer picture of trade-offs across time horizons, health states and standard and reverse iterations. Importantly, this study was only the second to evaluate how search procedures affect standard and reverse TTO values. As such, several important contributions were made in terms of the validity of the search procedures and a number of interesting observations emerged providing possible directions for future research. For one, sensitivity to health state severity appears to be more acute in choice search procedures than in matching procedures. The inconsistencies shown between search procedures and strategically equivalent variants of the TTO in this study jeopardize the TTO’s usefulness as a method of eliciting health state values. Moving forward, further investigations as to why these value differentials exist should help to identify the conditions under which the TTO is most useful and/or the means by which the various influences may be accounted for in TTO values.

References

Abhyankar, P., Volk, R. J., Blumenthal-Barby, J., Bravo, P., Buchholz, A., Ozanne, E., . . . Stalmeier, P. (2013). Balancing the presentation of information and options in patient decision aids: an updated review. *BMC Medical Informatics and Decision Making, 13*(Suppl 2), S6.

Arnesen, T., & Trommald, M. (2005). Are QALYs based on time trade-off comparable? - A systematic review of TTO methodologies. *Health Economics, 14*, 39-53.

Attema, A. E. (2012). Developments in time preference and their implications for medical decision making. *Journal of the Operational Research Society, 63*(10), 1388-1399.

Attema, A. E., & Brouwer, W. B. (2008). Can we fix it? Yes we can! But what? A new test of procedural invariance in TTO- measurement. *Health Economics, 17*(7), 877-885.

Attema, A. E., & Brouwer, W. B. (2010). The value of correcting values: influence and importance of correcting TTO scores for time preference. *Value in Health, 13*(8), 879-884.

Attema, A. E., & Brouwer, W. B. (2012). Constantly proving the opposite? A test of CPTO using a broad time horizon and correcting for discounting. *Quality of Life Research, 21*(1), 25-34.

Attema, A. E., & Brouwer, W. B. (2012). The way that you do it? An elaborate test of procedural invariance of TTO, using a choice-based design. *The European Journal of Health Economics*, *13*(4), 491-500.

Attema, A. E. and Brouwer, W. B.. (2013). In Search of a Preferred Preference Elicitation Method: A Test of the Internal Consistency of Choice and Matching Tasks. *Journal of Economic Psychology* 39 (Complete): 126-140. doi:10.1016/j.joep.2013.07.009.

Bachman, J. G., & O'Malley, P. M. (1984). Yea-Saying, Nay-Saying, and Going to Extremes: Black-White Differences in Response Styles. *Public Opinion Quarterly, 48*(2), 491-509. doi: 10.1086/268845

Bleichrodt, H. (2002). A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Economics, 11*, 447-456.

Bleichrodt, H., Luis Pinto, J., & Abellan-Perpinan, J. M. (2003). A consistency test of the time trade-off. *Journal of Health Economics, 22*, 1037-1052.

Bleichrodt, H., & Pinto, J. L. (2002). Loss Aversion and Scale Compatibility in Two-Attribute Trade-Offs. *Journal of Mathematical Psychology, 46*, 315–337.

Bostic, R., Herrnstein, R. J., & Luce, R. D. (1990). The effect on the preference-reversal phenomenon of using choice indifferences. *Journal of Economic Behavior & Organization,* *13*(2), 193-212.

Boye, K. S., Matza, L. S., Feeny, D. H., Johnston, J. A., Bowman, L., & Jordan, J. B. (2014). Challenges to time trade-off utility assessment methods: when should you consider alternative approaches? *Expert Rev Pharmacoecon Outcomes Res, 14*(3), 437-450. doi: 10.1586/14737167.2014.912562

Cairns, J. (1992). Discounting and health benefits: another perspective. *Health Econ, 1*(1), 76-79.

Dolan, P., & Gudex, C. (1995). Time preference, duration and health state valuations. *Health Economics, 4*(4), 289-299.

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

Dolan, P., & Roberts, J. (2002). To what extent can we explain time trade-off values from other information about respondents? *Soc Sci Med, 54*(6), 919-929.

Dolan, P., & Stalmeier, P. (2003). The validity of time trade-off values in calculating QALYs: constant proportional time trade-off versus the proportional heuristic. *Journal of Health Economics, 22*(3), 445-458.

Essink-Bot, M.-L., Stuifbergen, M., Meerding, W.-J., Looman, C., Bonsel, G., & Group, V. (2007). Individual differences in the use of the response scale determine valuations of hypothetical health states: an empirical study. *BioMed Central Health Services Research, 7*(62).

Heintz, E., Krol, M., & Levin, L.-Å. (2013). The Impact of Patients' Subjective Life Expectancy on Time Tradeoff Valuations. *Medical Decision Making, 33*(2), 261-270.

Hsee, C. K. (1996). Elastic justification: How unjustifiable factors influence judgments. *Organizational Behavior and Human Decision Processes, 66*(1), 122-129.

Jansen, S. J., Stiggelbout, A. M., Wakker, P. P., Vliet Vlieland, T. P., Leer, J. W., Nooy, M. A., & Kievit, J. (1998). Patients' utilities for cancer treatments: a study of the chained procedure for the standard gamble and time tradeoff. *Med Decis Making, 18*(4), 391-399.

Keeney, R. L., Von Winterfeldt, D., & Eppel, T. (1990). Eliciting public values for complex policy decisions. *Management Science, 36*(9), 1011-1030.

Knetsch, J. L., & Sinden, J. A. (1984). Willingness to pay and compensation demanded: experimental evidence of an unexpected disparity in measures of value. *The Quarterly Journal of Economics, 99*(3), 507-521. doi: 10.2307/1885962

Lenert, L. A., Cher, D. J., Goldstein, M. K., Bergen, M. R., & Garber, A. (1998). The effect of search procedures on utility elicitations. *Medical Decision Making, 18*(1), 76-83.

McNeil, B., Weichselbaum, R., & Pauker, S. (1981). Tradeoffs between quality and quantity of life in laryngeal cancer. *New England Journal of Medicine, 305*, 982–987.

Miyamoto, J., & Eraker, S. (1988). A multiplicative model of the utility of survival duration and health quality. *Journal of Experimental Psychology, 117*(1), 3-20.

Myerson, J., & Green, L. (1995). Discounting of delayed rewards: Models of individual choice. *Journal of the experimental analysis of behavior, 64*(3), 263-276.

Oliver, A. (2004). Testing the internal consistency of the standard gamble in 'success' and 'failure' frames. *Soc Sci Med, 58*(11), 2219-2229. doi: 10.1016/j.socscimed.2003.08.024

Oliver, A., & Wolff, J. (2014). Are people consistent when trading time for health? *Econ Hum Biol, 15c*, 41-46. doi: 10.1016/j.ehb.2014.05.001

Pliskin, J. S., Shepard, D. S., & Weinstein, M. C. (1980). Utility functions for life years and health status. *Operations Research, 28*(1), 206-224. doi: 10.1287/opre.28.1.206

Redelmeier, D. A., & Heller, D. N. (1993). Time preference in medical decision making and cost-effectiveness analysis. *Med Decis Making, 13*(3), 212-217.

Sackett, D., & Torrance, G. (1978). The utility of different health states as perceived by the general public. *Journal of Chronic Diseases, 31*(11), 697-704.

Schmidt, U., & Hey, J. D. (2004). Are preference reversals errors? An experimental investigation. *Journal of Risk and Uncertainty, 29*(3), 207-218.

Selart, M. (1996). Structure compatibility and restructuring in judgment and choice. *Organizational Behavior and Human Decision Processes, 65*(2), 106-116.

Selart, M. (1996). Structure compatibility and restructuring in judgment and choice. *Organizational Behavior and Human Decision Processes, 65*(2), 106-116.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3-4), 591-611. doi: 10.1093/biomet/52.3-4.591

Spencer, A. (2003). The TTO method and procedural invariance. *Health Economics, 12*(8), 655-668.

Stalmeier, P. F., Lamers, L. M., Busschbach, J. J., & Krabbe, P. F. (2007). On the assessment of preferences for health and duration: maximal endurable time and better than dead preferences. *Med Care, 45*(9), 835-841. doi: 10.1097/MLR.0b013e3180ca9ac5

Stiggelbout, A., Kiebert, G., Kievit, J., Leer, J., Habbema, J., & De Haes, J. (1995). The 'utility' of the time trade-off method in cancer patients: feasibility and proportional trade-off. *Journal of Clinical Epidemiology, 48*(10), 1207-1214.

Sutherland, H., Llewellyn-Thomas, H., Boyd, N., & Till, J. (1982). Attitudes towards survival. The concept of "maximum endurable time". *Medical Decision Making, 2*, 299-309.

Thaler, R. H. (1991). Some empirical evidence on dynamic inconsistency. *Quasi rational* *economics, 1*, 127-136.

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

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124-1131. doi: 10.1126/science.185.4157.1124

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453-458. doi: 10.1126/science.7455683

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics, 106*(4), 1039-1061.

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

van der Pol, M., & Roux, L. (2005). Time preference bias in time trade-off. *Eur J Health Econ, 6*(2), 107-111.

van Der Pol, M. M., & Cairns, J. A. (2000). Negative and zero time preference for health. *Health Economics, 9*(2), 171-175. doi: 10.1002/(SICI)1099-1050(200003)9:2<171::AID-HEC492>3.0.CO;2-Z

van Nooten, F., & Brouwer, B. (2004). The influence of subjective expectations about length and quality of life on time trade-off answers. *Health Economics, 13*, 819-823.

West, R. R., McNabb, R., Thompson, A. G., Sheldon, T. A., & Grimley Evans, J. (2003). Estimating implied rates of discount in healthcare decision-making. *Health Technol Assess, 7*(38), 1-60.

Wittenberg, E., Halpern, E., Divi, N., Prosser, L. A., Araki, S. S., & Weeks, J. C. (2006). The effect of age, race and gender on preference scores for hypothetical health states. *Quality of Life Research, 15*(4), 645-653.

End of Chapter 4 Notes:

i. In theory, discounting should not affect standard and reverse TTO values – i.e. if invariance holds. However, if procedural invariance is violated, then discounting cannot be assumed to exert a similar effect on both standard and reverse TTO values, since, for example, the time traded off in the standard TTO may affected by loss aversion, resulting in a longer number of years in full health which are to be discounted.

ii. Note, the choice search procedure used by Bleichrodt et al. (2003) differs from the choice-based procedures used in the other studies in Table 4-2 in that instead of showing respondents each choice scenario in isolation, the respondents were shown all of the choices they were to make in the valuation procedure at the same time. The ability to see all possible options simultaneously may have influenced decision processes used to arrive at their indifference value and may perhaps be more reflective of a matching procedure insofar as respondents are aware of all possible trade-off scenarios at once.

iii. Strictly speaking, the TTO requires the respondent to state the number of years they require in full health (standard TTO) or in the deteriorated health state (reverse TTO) in order to achieve indifference with the fixed outcome. To the same extent, the respondent may be asked to provide the number of years at which point they marginally prefer one outcome over another and assumed that values are likely to be contained within a reasonable margin of the range of indifferences (Oliver, 2004). Eliciting marginal preference has been reported as being more easily understood by respondents in this study and in others (e.g. Oliver, 2004).

iv. A possible discrepancy in the description of the back pain health state was identified on review of the draft of this manuscript. Specifically, the health state is described as frequent back pain in each of the narrative statements in the scenario but as 'no pain or discomfort' in the EQ-5D dimensions. The EQ5D description of back pain was drawn from Attema and Brouwer (2013). If respondents perceived a discrepancy between back pain in the narrative and the absence of pain in the EQ5D dimensions, they may have interpreted the health state as less severe than intended (i.e. focusing only on their limitation in terms of walking about). This should not have impacted the within-BP comparisons since all respondents saw the same definition for back pain. Further, comparisons between back pain and unwell values should not have been influenced since respondents would have interpreted the back pain health state either as intended (i.e. with both limitations in walking about and with some pain), or as a less severe health state. A second possibility is that there may have been variability among respondents in how they interpreted the health state and the resulting values. Increased variability in terms of the interpretation of the health state might have masked differences that would otherwise have been apparent; however, differences were statistically significant in key comparisons in terms the study's objectives. That respondents interpreted the back pain health state as intended is suggested by the finding of variation in values similar to other reports and that no respondents remarked that they had noticed this discrepancy.

v. Congruent with the non-traders in Spencer (2003) and Oliver and Wolff (2014), all of the respondents who were unwilling to trade in the reverse TTO considered the health state to be better than death in the standard TTO (i.e. below a certain length of life they would opt to live longer in the worse health state), highlighting an inconsistency between the two TTO constructs (standard and reverse) in this respect.

Chapter 5. Discussion – Broader Implications of the Research

Decision makers are faced with increasing innovation and demand, accompanied by escalating costs. Thus, as we are living longer than ever, but not always better, institutions and governments have embraced initiatives aimed at promoting health care value (i.e. better outcomes per moneys spent). Value-based health care (VBH), in particular, has risen to the forefront of policy agendas. The underlying premise of VBH is that the prices paid for new technologies should be reflective of their overall value (Eldessouki and Dix Smith, 2012). A prerequisite for VBH is the identification of affordable and effective technologies, and thus to provide effective health care, we must first come to a consensus on defining the parameters of value, including its composite components – i.e. health and, potentially, non-health outcomes).

A summary measure of HRQoL, such as the QALY, to help decide how to allocate available resources is therefore more desirable than ever. In no other area of public policy has a measure similar to the QALY been developed, and the QALY is unique therefore in both its ambitions and in the political, philosophical and measurement challenges which it faces.

This thesis set out to examine health state valuation using the TTO in the context of a behavioural economic framework. Procedural and descriptive invariance are cornerstones of decision theory – on which current health state valuation techniques depend for valid input into economic analysis – yet violations of these basic assumptions are widely observed in health state valuation and elsewhere. Behavioural economics offers a framework by which such inconsistencies can potentially be better understood. Although behavioural economic concepts have gained traction in decision research in other disciplines, its application to health state valuation remains nascent.

Three empirical studies focused on key problems in the measurement and valuation of HRQoL using the TTO. Below, the main conclusions are presented and discussed, limitations of the research presented in the thesis are acknowledged, and challenges for future investigations/tors are highlighted.

5.1 The Studies

Chapter 2: The study presented in Chapter 2 examines a fundamental question which for the most part has been taken for granted in TTO research to date – whether respondents are able to adequately consider the trade-off between longevity and HRQoL within the context of their own life. This study examined the effect of using an age-framed TTO, whereby the time horizon is expressed as the respondent's age at death (calculated by adding their current age to the time horizon in question). It was observed that when TTO values were elicited using a matching search procedure, no differences emerged between the age- and standard frames but when values were elicited using a bisection search procedure, age-frame values were actually *lower* than those elicited through the standard TTO. The results, therefore, were not entirely supportive of the initial hypothesis that framing the time horizon in terms of the respondent's age at death would result in greater loss aversion, and thus higher values, than the standard TTO. The qualitative findings provided insights into what factors the respondents consider when making trade-offs and, notably, many respondents provided rationales for their trade-offs that deviate from the intended compensatory relationship between HRQoL and longevity.

While it has been recognized that the TTO is a somewhat abstract exercise, little research has sought to understand the degree to which respondents appropriately interpret the task and, particularly, project the provided situation onto their own lives. A further issue arises in terms of comparing the valuations of respondents who may interpret the task in different ways and whether the particular construction of the TTO (e.g. the time horizon used) influences the accurate interpretation of the task. Moving forward, it may be fruitful to for researchers to gain a better understanding as to why age-framed valuations differ from standard valuations since it could be argued that the age-framed scenario is of greater realism and therefore potentially capable of yielding more accurate valuations than the standard frame, though this remains uncertain.

Chapter 3: The aim of the study presented in Chapter 3 was to examine the TTO values of younger and older respondents when varying the time horizon. It was expected that, due to the behavioural economic principle of loss aversion, younger respondents would be less likely to trade longevity for HRQoL at shorter durations compared to older respondents. The study also investigated the effects of framing the TTO time horizon in term of the respondent's perceived life expectancy. In implementing a life expectancy time horizon, it was expected that younger respondents would have lower TTO values than older respondents due to decreasing marginal utility for time and thus a greater willingness to forgo years further in to the future.

The findings showed that when presented with TTOs of varied time horizons, respondents appeared to adopt different decision strategies depending on the length of the time horizon. In particular, the results suggest that when respondents faced a short time horizon (unrealistically short in the case of younger respondents), they adopted decision heuristics to arrive at their responses. More specifically, qualitative evidence suggested that respondents rely on rules of thumb – particularly, proportionality (e.g. ½ or ¾ of the time horizon) – as a means of arriving at their trade-off decisions. Using a life expectancy time horizon presents several advantages among which, intuitively, is that the respondent is presented with a more realistic scenario. Matza et al. (2015) have proposed that a life expectancy time horizon enabled respondents to focus more on the health state in question. Additionally, Dave et al. (2010) as in Attema and Brouwer (2012, p. 423) note that existing research "suggests there is a tradeoff between predictive accuracy on the one hand, and cognitive limitations on the other." As such, future research would benefit from probing the relationship between respondent age, the time horizon, and TTO values.

Chapter 4: A number of hypotheses were tested in the third study. Of main interest was the performance of the reverse TTO relative to the standard TTO construction and the question of procedural invariance. Whether the standard and reverse TTO elicit the same health state values or not is important to address, not least because one version or the other (standard or reverse) may present a closer approximation of, and be of greater relevance, to real-world decision-making contexts. This may, for example, be especially useful for addressing challenges that arise in the context of end-of-life decisions found when using existing elicitation methods and when high cost/lower return scenarios are pertinent.

A limited number of studies have shown that reverse TTO values are lower than standard TTO values due to loss aversion (which increases standard TTO values). The design was intended to either confirm or refute published findings and to build on them by incorporating additional tests, thus forming meaningful links between earlier research and the work presented in this thesis. Behavioural effects may differ by search procedure and the magnitude of such effects may depend on other aspects of the TTOs construction (Versteegh et al., 2013). This study also included a secondary question relating to the effect of health state severity.

The study found evidence of procedural invariance in two forms, both between standard and reverse TTO values as well as between values elicited using bisection and matching. Inconsistencies were more frequently observed when a more severe health state was considered. An additional analysis showed that a matching search procedure yielded valuations that were comparable for a relatively mild and severe health state, thus appearing to be insufficiently responsive to capturing changes in health state severity. Matching is a frequently relied upon method for eliciting preferences, both in health state valuation and elsewhere, since it offers a compromise between the effort asked of respondents (i.e. number of questions asked) and the responses obtained (Deparis et al., 2015). This result therefore suggests that researchers implementing this search procedure should be cautious of whether the compromise between respondent fatigue and number of questions answered is in fact worthwhile.

The strict focus on TTO throughout this thesis was intended, as the TTO is the most widely implemented tool used in the elicitation of health state preferences both in a research context and in clinical studies which seek to demonstrate cost-effectiveness (Dolan et al., 1996). However, the empirical findings and discussion in this thesis are of relevance not only to researchers of health state valuation but also to policy makers in health and other areas of policy and governance which seek to inform decisions (e.g. through policy evaluation) with the help of stated preference exercises.

5.2 Web-based Survey Design

The study described in Chapter 2 used a web-based survey design which presents a number of advantages as well as potential challenges when compared to face-to-face interviews. Despite some unanswered questions, web-based research appears to be a useful resource for methodological researchers and has been growing in recent years, particularly among researchers of TTO methodology.

For the TTO, wed-based surveys offer a useful means by which particular items or questionnaires in their entirety can be quickly iterated and tested at relatively lower costs to the researcher (Oppe et al., 2016). To this extent, web-based formats may serve as convenient grounds for piloting surveys. Web-based surveys also have a potential advantage in terms of the recruitment of respondents from a greater radius than face-to-face interviews would permit, therefore possibly allowing for a more diverse respondent pool and the opportunity to access target groups who might otherwise be difficult to reach and engage in research. As discussed in Chapter 2, web-based platforms have been successful in replicating economic and psychology experiments, similar to those undertaken in this thesis (see references in Chapter 2). Chapter 2 further adds to the existing literature on self-complete health state valuation methods by showing that this method offers a useful alternative to face-to-face interviews while also collection rich qualitative data. Finally, although dependent on the respondent, online self-complete surveys have the benefit in that respondents can complete the questionnaire at their own convenience which may lead to better engagement levels. Importantly, however, the converse may also hold true – in the absence of a face-to-face interviewer, respondents may be more prone to distraction or speeding through responses

There are clearly unresolved questions with regards to using web-based surveys. Perhaps the most salient question surrounds the unknown level of respondent comprehension and engagement in the task and the effort they invest in each question. To this end, there is evidence that respondents who are less invested in the task may provide responses that are more affected by behavioural influences. For example, Augestad et al. (2016) suggest that cognitive effort could be tied to the anchoring and adjustment bias, whereby the respondent anchors to an outcome and fails to adequately adjust away from this outcome in light of new information (Kahneman and Tversky, 1974).

Oppe et al. (2016) comment that in comparisons of web-based versus face-to-face interviews, respondents who completed the former version reached their point of indifference after fewer iteration steps (in a bisection procedure) than did respondents in the face-to-face condition. While such findings may be indicative of poorer engagement in an online setting, it is worth noting that these types of engagement issues might be attenuated by informing the respondent that repeated trade-offs/questions of a similar type will be presented to them and that they should carefully consider each trade-off in its own right. In terms of challenges related to sampling, online research may exclude particular subgroups (e.g. those with limited access to computer or web) (Craig et al., 2014). It is worth considering that any method chosen will to varying degrees exclude some respondents. In the case of face-to-face interviews, respondents who are unable to attend at available times or travel will likely be left out of the sample. An additional difficulty raised with online sampling is that it is difficult to confirm the demographics stated by the respondent. Some researchers may wish to use authentication strategies such as requiring photographic or other means of identification to corroborate the respondents' demographic details, for example; however, it is unknown how widely this strategy may be implemented.

It is worth commenting that there are other possibilities which may leverage the benefits of both online and in-person interviews, such as remote interviews via webconference, these techniques have been relatively less explored in health state valuation research. Further, some researchers have adopted iterations of face-to-face interviews implementing CAPI (computer assisted personal interviewing) interviews to multiple respondents at a time in an in-person setting to reap certain benefits of web-based platforms while also encouraging comprehension and accountability by having the respondent in the presence of the researcher.

While there are certainly benefits and disadvantages to both methods discussed here (Internet/web-based survey and face-to-face interviews), it seems likely that the chosen method will depend largely on the purpose of the research. That is, for studies of methodological aspects of the TTO, researchers may show increasing reliance on web-based technologies and until greater validity of such methods is established, international research programmes (i.e. that of the EuroQoL group) will indefinitely continue to conduct face-to-face interviews.

5.3 Limitations to the Research

There are a number of limitations to the research presented in this thesis. One potential source of criticism from economists is whether the respondents had sufficient incentive to provide their truthful valuation. In all three studies, trade-offs were hypothetical and thus the results may be susceptible to hypothetical bias (i.e. the difference in response or preference which results from it having been elicited in a hypothetical elicitation versus observed through revealed preferences). Health state valuation exercises are typically undertaken in the absence of incentives for a number of reasons, primarily the difficulty in assigning a monetary amount to the outcomes under consideration and the inability, let alone undesirability, of recreating real-world circumstances.

Bleichrodt and Pinto (2002) have commented that the hypothetical nature of their health state value elicitation tasks is not problematic in terms of respondents' efforts to provide 'truthful' answers. In support of this conjecture, a number of empirical investigations have shown no systematic differences in responses elicited between hypothetical and incentivized scenarios (Beattie and Loomes, 1997; Camerer, 1995; Camerer and Hogarth, 1999; Tversky and Kahneman, 1992). Munro (2009), for instance, demonstrated that deviations from rational decision-making as defined by standard economic theory are pervasive and apply to experiments that have and have not implemented economic incentives (notably, also in real-world field experiments).

A second limitation concerns the generalizability of the results. In Chapters 3 and 4, for example, convenience samples consisting mainly of higher education students were used. These samples were therefore rather young and highly-educated and it is possible that another sample with different characteristics would yield different findings. That being said, respondents were not recruited with the aim of obtaining samples reflective of the general population (e.g. as might be desirable in policy decisions) but rather to conduct a series of tests assessing the stamina of the QALY model as it applies to the TTO elicitation method.

5.4 Fundamental Issues and Policy Implications

Although there has been substantial discussion among philosophers and economists on ways of improving particular tools of quantification, there is far less work on the fundamental question of *whether to quantify at all*. Although the notion of value-weighted time underlying the TTO is clearly desirable, based on empirical findings, it would be imprudent not to seriously question whether we have leapt from the pragmatic to the applied all too quickly. Arnesen and Norheim (2003, p. 84) note, "the great uses one might make of quality of life weightings *were they available* does not, however, imply that it is possible to elicit them." As shown in this research and elsewhere, it is debatable that conceptually, meaningful preferences for health states can be elicited using the TTO (i.e. by simply trading longevity for HRQoL). Surprisingly, there is very little research into the fundamental issues of the TTO.

In line with many of the findings in this thesis, Schwartz (2006) remarks that when faced with medical decisions, individuals consider not only how long they might live and in what state of health but in addition what they are living for (e.g. children) and under what conditions (i.e. in what health state) they would be willing to sacrifice years of their life. This means the consequences considered by respondents in their TTO decisions often extend far beyond those taken into account by standard models. It has also been proposed that significant differences exist between conditions that individuals *fear* compared to those which have meaningful impact on HRQoL (Dolan, 2007). It follows that if respondents are providing TTO responses which are based on fear, then resources may be directed in such a manner as to reduce fear as opposed to improvements in HRQoL.

Questions relating to morbidity and mortality are incredibly complex and therefore the assumptions that they can be meaningfully reduced to a single numerical value and that trading HRQoL for longevity can capture an accurate valuation of a health state would, intuitively, seem far-reaching (Samuelsen, 2011; Oliver, 2013). Moreover, the inseparability between length of life and quality of life, such that length of life must necessarily accompany HRQoL raises conceptual problems.

5.5 Other directions

Over the past 30 years, the conceptual foundation of the QALY has remained the same. Despite this, the QALY and the methods used to derive HRQoL have not been superseded by another measure (Cuyler and Kobelt, 2014). In response, suggestions have been put forward that perhaps we should seek to quantify and optimize something other than HRQoL.

Perhaps most prominently, some researchers and policy-makers have become interested in *happiness research* (Dolan, 2011, p.3) and the evaluation of subjective well-being (SWB) as an alternative to attempting HRQoL valuations and as a public policy objective (Layard, 2005). In brief, Brazier and Tsuchiya (2015) describe well-being as "how well an individual's life is going," while the literature describes subjective well-being, for instance, in terms of hedonism (i.e. an individual's well-being increases when they experience pleasure and decreases when they experience pain) (Hirschauer et al., 2015).

The measurement of SWB is advocated by commentators who contend that public policy might best be guided through experience utility as opposed to preference-based decision utility – the latter of which is the case with using the general population in TTO elicitations (e.g. Kahneman et al., 1997; Dolan and Kahneman, 2008). This school of thought has gained many advocates and followers in recent years who support the notion that maximizing happiness (e.g. Layard, 2005) or wellbeing should be the goal of public policy.

Other suggestions have involved alternatives such as multiple criteria decision-making analysis (MCDA) or discrete choice experiments (DCE) (Hansen, 2012; Oppe et al., 2014; Lancsar and Louviere, 2008). MCDA, for instance, include health-related considerations in addition to other non-health outcomes (e.g. *process* characteristics) (Mooney, 1994). It is important to recognize, however, that alternative methods, including MCDA and DCE, are faced with their own challenges and it is difficult to know when to pursue the 'perfect' at the expense of finding (or refining) the 'good' (i.e. what is already in use, the QALY).

As the examples above continue to rely to varying degrees on individual preferences, an important question is whether it is morally defensible to guide policy decisions if individuals are not always aware of what is in their best interest and, as a result, decision-makers risk informing their decisions on potentially biased preferences (values) (Menzel, 2014).

Recognizing that individuals may not be in an adequate position to undertake such decisions or even possess the desire to do so raises the issue of whether to maintain preference (consumer) sovereignty should be upheld in terms of health state valuation. Stepping away from the use of individual preferences to gauge the value of health interventions, a possible option is to use a panel of experts to provide cost-effectiveness determinations. Those who argue in favour of using expert opinion as a means of allocating resources suggest that policy formation or resource allocation be performed by expert informants in both the area of interest alongside experienced policy officials (Peterson, 2001). Recognizing that such a technocratic approach, in that the general population is relieved of its role in helping to 'value health', may be unappealing to some, however, who interpret this approach as paternalistic. In defence of such a position, some research has shown that instead of making their own decisions, individuals may seek to defer health care decision to a loved one or relative, or to someone with expert knowledge or greater experience (e.g. physician). Chen et al. (2011) comment that while there is limited research on choice deferral, when decisions involve trade-offs between highly valued attributed (such is the case with longevity and HRQoL), choice deferral is more common (Beattie and Barlas, 2001; Luce, 1998).

Considering the limitations imposed by value inconsistencies, some have advocated for the adoption of a single, standardized methodology. Developing standardized methods aims to address a key issue surrounding the potential loss of comparability between TTO values that arise from using different time horizons and search procedures, among other variables (Prosser et al., 2012). The idea behind using the same methods to elicit preferences is that although biases may be present, they can be assumed to be consistent. Suppose, for instance, if the TTO was to consistently provide upwardly biased estimates. If values have been elicited using the same methodology, they can therefore be assumed to be comparable (with the further assumptions that biases are both in the same direction and magnitude) (Oliver and Wolff, 2014). As such, standardized methods, such as the MVH approach used in EQ5D valuations have been adopted in a wide range of studies (Szende et al., 2007). Checklists and reference cases have also been put forward in attempts to streamline implementation and flag potential issues of comparability. Attema et al. (2013) and Versteegh et al. (2013) devised a checklist with methodological specifications, enabling researchers to align their methodologies with one another and improve the comparability of resulting TTO values.

Despite efforts promoting standardization, in practice, TTO methods elicited using different methods are often compared and therefore the extent to which methodological inconsistencies in CUA may ultimately affect policy decisions remains unknown (McDonough and Tosteson, 2007). Matza et al. (2015) comment, specifically with reference to the time horizon, that studies often don't provide justification for their methodological choices.

Alternatively, from an industry standpoint, given the known issues of comparability of TTO values elicited using different versions of the same tool, inconsistencies leave open the possibility of misuse (Joore et al., 2010). For example, given that it is in the best interest of pharmaceutical companies to have interventions that are cost-effective, companies may look to 'cherry-pick' more favourable valuation results for submissions to reimbursement agencies. On the other hand, companies may face challenges insofar as yielding accurate forecast predictions, for example, when attempting to determine a drug's pre-market value.

Preferable methods, however, are challenging to define. Yet, ultimately, standardization may be at the expense of the validity of the resulting health state value. Specifically, in respect of inconsistent values yielded from different time horizons, Buckingham and Devlin (2009, p. 364) remarked:

…the use of a single 'tariff' of TTO values for health states, such as that used by NICE (2004) and other similar decision-making bodies internationally, will yield estimates of QALYs that are, in many cases, quite simply wrong.

5.6 Conclusion

A fundamental objective of healthcare systems is to strike the best possible balance between providing limited resources both efficiently and equitably. Economic evaluation is referred to as a tool to help inform policy makers of the relative value of different choices. A critical question remains of whether health state valuation tools, including the TTO, are able to adequately capture HRQoL. As long as QALYs continue to be used for input into economic evaluation, the extent to which the limitations of HRQoL elicitation tools compromise the value of such inputs remains an issue of great merit.

Brazier et al. (2007, p. 118) commented:

> … a decision maker needs a common and consistent set of methods for informing decisions and so choices need to be made about technique, variant and course of values. Such choices are made for reasons of expediency for assisting decision making … but it is important to continue research into major outstanding issues such as the theoretical basis of different techniques, ways to improve their reliability and validity (for a given concept).

Drawing upon decision-making literature across a number of disciplines, this thesis provides evidence that health state values are influenced to an important extent by the construction of the TTO elicitation exercise, that some variants of TTO construction are more susceptible to violations of procedural and descriptive invariances than others, that some variants are more sensitive to differences in health states than others, that some variants may be more likely to evoke heuristics and other non-compensatory decision strategies, and that many of these findings can be accounted for within a behavioural economics framework.

The thesis presents a range of examples which showcase violations of internal consistency, particularly relating to respondent age and the time horizon, search procedures and standard and reverse TTO constructions. Some of these inconsistencies have been explored in the existing literature limited, though in a piecemeal fashion. The thesis generated a number of insights into respondents’ cognitive processes and how they may be affected by the TTO design (including compensatory and non-compensatory decision-making strategies), a relatively unexplored area in the TTO literature. In addition, the studies have provided novel observations about choice/matching responsiveness to differences in health state severity and added to the literature surrounding the increasingly popular use of web platforms for health state valuation.

Taken together, the empirical findings presented in this thesis offer clear support for the view that TTO values do not fully abide by the decision theoretic principles upon which their implementation in economic evaluation is based. The choice of the specific elicitation procedure used in stated preference exercises can have significant influence on the reported preferences and thus on priority setting and the allocation of public resources. These results are consistent with a well-developed literature in other domains investigating individual decision-making and a growing body of evidence in health state elicitation.

What remains relatively unknown, and where this research aimed to provide clarity, is the extent to which different constructions of the TTO reflect preferences for different health states (Hazen, 2006; Robinson and Hammitt, 2010).

A substantial amount of investment and intellectual capital power has been invested in using the QALY both in practical applications in addition to academic/policy research. Regardless of whether or not the QALY continues to exert the same level of influence on health policy that it does today, it is important to recognize that [these investments] have provided key insights contributing to the broader decision making literature, in particular how the presentation of health-specific data is received by individuals who rarely face such decisions. Botzen (2011) explains that 'unresolvedness' in research is not necessarily detrimental to its value, but rather may reflect the current state of affairs in health state preference elicitation, more specifically deviations of the TTO from its underlying model. To this extent, the overarching contribution of this thesis is its extension on the limited body of literature assessing validity of TTO values as well as the underlying assumptions of the TTO model. In efforts to address a range of research questions throughout, the thesis has also brought forward a number of additional issues which should be considered in future research in this area: How might valuation exercises be presented so as to be more realistic to respondents, thereby likely minimizing the influence of behavioural influences, and, more fundamentally, whether the QALY, as it stands is a suitable "machinery for making social choices from individual tastes" (Haubrich and Wolff, 2009, p. 10).

Policy-making is an inherently complex exercise. It is clear that relying on stated preference methods such as the TTO to assign value to health interventions is an imperfect strategy, leading critics of the QALY to argue that it should be abandoned in favour of other approaches. Nonetheless, many would agree that to date no superior metric or strategy has emerged as preferable [to the QALY] and that therefore it remains "an indispensable tool" in health care priority setting (Smith et al., 2009). As such, if the TTO is to remain a cornerstone of health state preference elicitation, we must seek to better define the 'cognitive architecture of decision making' (Mehta, 2013).

References

Arnesen, T. M. and O. F. Norheim (2003). "Quantifying quality of life for economic analysis: time out for time tradeoff." *Med Humanit* **29**(2): 81-86.

Attema, A. and B. Brouwer (2008). "Can we fix it? Yes we can! But what? A new test of procedural invariance in TTO-measurement." *Health Economics* **17**: 877-885.

Attema, A. and B. Brouwer (2010). "On the (not so) constant proportional trade-off in TTO." *Quality of Life Research* **19**: 489-497.

Attema, A. E. and W. B. F. Brouwer (2012). "Constantly proving the opposite? A test of CPTO using a broad time horizon and correcting for discounting." *Quality of Life Research* **21**(1): 25-34.

Attema, A. E. and W. B. F. Brouwer (2012). "The way that you do it? An elaborate test of procedural invariance of TTO, using a choice-based design." *The European Journal of Health Economics* **13**(4): 491-500.

Attema, A. E., et al. (2013). "Time trade-off: one methodology, different methods." The *European Journal of Health Economics* **14**(Suppl 1): 53-64.

Attema, A. E., et al. (2013). "Lead time TTO: leading to better health state valuations?" *Health Economics* **22**(4): 376-392.

Augestad, L. A., et al. (2016). "Influenced from the start: anchoring bias in time trade-off valuations." *Qual Life Res* **25**(9): 2179-2191.

Beattie J, Barlas S. *Predicting perceived differences in tradeoff difficulty*. In: Weber E, Baron J, Loomes G, editors. Conflict and tradeoffs in decision making. Cambridge University Press; New York, NY: 2001. pp. 25–64.

Beattie, J. and G. Loomes (1997). "The impact of incentives upon risky choice experiments." *Journal of Risk and Uncertainty* **14**(2): 155-168.

Beresniak, A., et al. (2015). "Validation of the underlying assumptions of the quality-adjusted life-years outcome: results from the ECHOUTCOME European project." *Pharmacoeconomics* **33**(1): 61-69.

Bleichrodt, H. and J. L. Pinto (2002). "Loss Aversion and Scale Compatibility in Two-Attribute Trade-Offs." *Journal of Mathematical Psychology* **46**: 315–337.

Botzen, W. J. (2011). "Alistair Munro: Bounded Rationality and Public Policy: A Perspective from Behavioural Economics. Ian J. Bateman (ed.): The Economics of

Non-Market Goods and Resources." *Environmental and Resource Economics* **49**(2): 305-308.

Brazier, J., et al. (2007). Measuring and Valuing Health Benefits for Economic Evaluation, OUP Oxford.

Brazier, J. and A. Tsuchiya (2015). "Improving Cross-Sector Comparisons: Going Beyond the Health-Related QALY." *Appl Health Econ Health Policy* **13**(6): 557-565.

Buckingham, K. J. and N. J. Devlin (2009). "A note on the nature of utility in time and health and implications for cost utility analysis." *Soc Sci Med* **68**(2): 362-367.

Camerer, Colin (1995) "Individual Decision Making", in Kagel, J. and A. E. Roth, Handbook of Experimental Economics, Princeton: Princeton University Press.

Camerer, C. F. and R. M. Hogarth (1999). "The effects of financial incentives in experiments: A review and capital-labor-production framework." *Journal of Risk and Uncertainty* **19**(1-3): 7-42.

Chen, Y., et al. (2011). "Age Differences in Trade-Off Decisions: Older Adults Prefer Choice Deferral." *Psychology and Aging* **26**(2): 269-273.

Cookson, R., & Culyer, A. (2010). Measuring overall population health: the use and abuse of QALYs. In A. Killoran (Ed.), Evidence-Based Public Health: Effectiveness and Efficiency, p. 148. Oxford University Press.

Craig, B. M., Reeve, B. B., Brown, P. M., Cella, D., Hays, R. D., Lipscomb, J., . . . Revicki, D. A. (2014). US Valuation of Health Outcomes Measured Using the PROMIS-29. *Value in Health*, 17(8), 846-853. doi: http://dx.doi.org/10.1016/j.jval.2014.09.005

Culyer, A.J. and Kobelt, G. eds. (2014) *Portrait of a health economist: essays by colleagues and friends of Bengt Jonsson.* Lund: Swedish Institute for Health Economics, pp.179-188

Dakin, H., et al. (2014). "The influence of cost-effectiveness and other factors on NICE decisions." *Health Economics*. doi: 10.1002/hec.3086

Deparis, S., et al. (2015). "The effect of bi-criteria conflict on matching-elicited preferences." *European Journal of Operational Research* **242**(3): 951-959.

Devlin, N. J., et al. (2011). "A uniform time trade off method for states better and worse than dead: feasibility study of the 'lead time' approach." *Health Econ* **20**(3): 348-361.

Dolan, P. (2007). *Finding a NICEr way* to value *health*: From hypothetical preferences to real experiences. SMF Foresight, The Social Market Foundation.

Dolan, P. (2011). "Thinking about it: thoughts about health and valuing QALYs." *Health Econ* **20**(12): 1407-1416.

Dolan, P., Gudex, C., Kind, P. and Williams, A. (1996), The time trade-off method: Results from a general population study. *Health Econ*., 5: 141–154. doi:10.1002/(SICI)1099-1050(199603)5:2<141::AID-HEC189>3.0.CO;2-N

Dolan, P. and D. Kahneman (2008). "Interpretations of utility and their implications for the valuation of health." *The Economic Journal* **118**: 215-234.

Eldessouki, R. and M. Dix Smith "Health Care System Information Sharing: A Step Toward Better Health Globally." *Value in Health Regional Issues* **1**(1): 118-120.

Hazen, G. (2007). "Adding Extrinsic Goals to the Quality-Adjusted Life Year Model." *Decision Analysis* **4**(1): 3-16.

Joore, M., et al. (2010). "The impact of differences in EQ-5D and SF-6D utility scores on the acceptability of cost-utility ratios: results across five trial-based cost-utility studies." *Value Health* **13**(2): 222-229.

Kahneman, D. and A. Tversky (1974). *Subjective probability: A judgment of representativeness*. The Concept of Probability in Psychological Experiments, Springer**:** 25-48.

Kahneman, D., et al. (1997). "Back to Bentham? Explorations of experienced utility." *The Quarterly Journal of Economics*: 375-405.

Lancsar, E., & Louviere, J. (2008). Conducting discrete choice experiments to inform healthcare decision making. *Pharmacoeconomics*, *26*(8), 661-677

Layard, R. (2005). *Happiness: Lessons from a New Science*, Penguin Press.

Luce MF. (1998) Choosing to avoid: Coping with negatively emotion-laden consumer decisions. *Journal of Consumer Research*: 24:409–433. doi:10.1086/209518.

Lundberg, L., et al. (1999). "Health-state utilities in a general population in relation to age, gender and socioeconomic factors." *European Journal of Public Health* **9**(3).

Matza, L. S., et al. (2015). "Acute and chronic impact of cardiovascular events on health state utilities." *BMC Health Services Research* **15**(1): 1-11.

McDonough, C. M. and A. N. Tosteson (2007). "Measuring preferences for cost-utility analysis." *Pharmacoeconomics* **25**(2): 93-106.

Mehta, J. (2013). "The discourse of bounded rationality in academic and policy arenas: pathologising the errant consumer." *Cambridge Journal of Economics*: bet040.

Menzel, P. T. (2014). *Utilities for health states: Whom to ask*. In Anthony J. Culyer (Ed.) Encyclopedia of Health Economics (5) 417-424. San Diego: Elsevier. doi 10.1016/B978-0-12-375678-7.00508-3.

Munro, A. (2009). *Bounded rationality and public policy: a perspective from behavioural economics,* Springer Science & Business Media.

National Institute for Health and Care Excellence. (2014). Value Based Assessment of Health Technologies. Consultation Paper. London, National Institute for Health and Care Excellence, Centre for Health Technology Evaluation.

Oliver, A. (2013). "From Nudging to Budging: Using Behavioural Economics to Inform Public Sector Policy." *Journal of Social Policy* **42**(04): 685-700.

Oliver, A. and J. Wolff (2014). "Are people consistent when trading time for health?" *Econ Hum Biol* **15**: 41-46.

Oppe, Mark, et al. (2014) "A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol." *Value in Health* 17.4: 445-453

Oppe, M., et al. (2016). "EuroQol Protocols for Time Trade-Off Valuation of Health Outcomes." *Pharmacoeconomics*.

Peterson, M. A. (2001). From Trust to Political Power: Interest Groups, Public Choice, and Health Care. *Journal of Health Politics, Policy and Law* 26(5), 1145-1163. Duke University Press. Retrieved August 30, 2016, from Project MUSE database.

Prosser, L. A., Grosse, S. D., & Wittenberg, E. (2012) Health utility elicitation: Is there still a role for direct methods? *Pharmacoeconomics* (2), 83-86

Robinson, L. and J. Hammitt (2011). "Behavioral Economics and the Conduct of Benefit-Cost Analysis: Towards Principles and Standards." *Journal of Benefit-Cost Analysis* **2**(2).

Samuelsen, C. H. (2011). *Anchoring in the lead-time trade-off: Does the starting point influence preference elicitation?* Master thesis. University of Oslo. Available at http://um.nb.no/URN:NBN:no-29703

Schwartz, S. H. (2006). "Basic human values: an overview." Recuperado de http://www.yourmorals.org/schwartz.

Smith, M. D., et al. (2009). "Moving the QALY forward: rationale for change." *Value in Health* **12**(s1): S1-S4.

Szende A, Oppe M, Devlin N. 2007. *EQ-5D Valuation Sets: An Inventory, Comparative Review and Users' Guide*. EuroQol Foundation. Springer: Rotterdam.

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

Versteegh, M. M., et al. (2013). "Time to tweak the TTO: results from a comparison of alternative specifications of the TTO." *The European Journal of Health Economics* **14**(Suppl 1): 43-51.

Wolff, J., & Haubrich, D. (2009-09-02). Economism and its Limits. *Oxford Handbooks Online*. Retrieved 10 Aug. 2016, from http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199548453.001.0001/oxfordhb-9780199548453-e-037.

List of Appendices

Appendix 2-1: Bisection Iteration Sequences

10-year TTO

The 10-year time horizon was initially bisected and then respondents were subsequently titrated upward or downward based on their responses. Respondents were asked to make trade-offs from 0.5+/- their most recent response. To illustrate, if the respondent had indicated that they would accept fewer than 3 years in full health but would not accept 2 years in full health, they were then asked to indicate their preference between 2.5 years in full health and *x* years in the inferior health state. The same procedure (i.e. choices to 0.5 units) was implemented in the 30 year TTO.

30-year TTO

Appendix 2-2: Discounting Task

Following a framework set out by Cairns and van der Pol (2000), respondents were asked to imagine that they would be ill (in the health state they were valuing in the TTO, i.e. either back pain or unwell) for a period of 25 days, 5 years in the future. Presented with the option of a treatment to delay the illness, they were then asked to state the number of days they would consider being ill (for the purpose of this example, *x* days) if they could postpone the illness so that it would occur 20 years in the future (West et al., 2003, termed the number of additional days they were willing to be ill for the 'delay premium') (as in Box A4-1, for example). Subsequently, respondents were asked to indicate the number of days they considered to be equal to being ill for *y* days in 20 years if they could postpone the illness to occur 35 years in the future. From this exercise, *r*, the respondent's discount rate can be calculated (as in Figure A4-1) and used to adjust uncorrected TTO values.

Box A4-1: Sample questions used to gauge respondents' implied discount rates

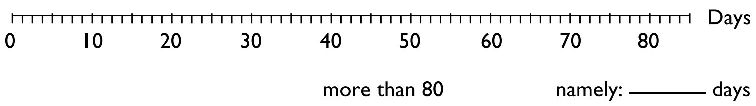
Imagine that you will be ill starting 3 years from now for 25 days. There is a treatment available that will postpone this period of ill health to a point further in the future. For instance, the treatment could have the following effects: your period of ill health would start 12 years from now instead of 3 years from now; and you would then be ill for 50 days instead of 25 days.

You might think this treatment is a good idea: the advantage of postponing the ill health outweighs the disadvantage of being ill for a longer period. Alternatively, you might think the treatment is not worthwhile: you do value the postponement but the advantage of this is outweighed by the disadvantage of being ill for a longer period (or you might simply prefer to be ill 3 years from now instead of 12 years from now).

Imagine that you will be ill [unwell/back pain] starting 5 years from now for 25 days and that treatment is available which will postpone this spell of ill health.

What is the maximum number of years of ill health that would still make the treatment worthwhile for you? For example, say that the treatment can postpone the period of ill health to 20 years in the future. If the number of days of ill health in that year were zero, probably everyone would choose the treatment. As the number of days of ill health in that year increases, individuals would at some point no longer prefer to be treated. What we are interested in is the maximum number of days of ill health at which you would still choose to be treated.

If the ill health [unwell/back pain] would then start 20 years from now, what is the maximum number of days of ill health that would still make the treatment worthwhile?



Box A4-2: Calculation of a respondent's implied discount rate

Question 1: 5 years from now 25 days ~ 20 years from now ***x*** days
Question 2: 20 years from now for ***x*** days ~ 35 years from now ***y*** days

Using the values of *x* and *y* provided by the respondents in Questions 1 and 2, the discount rate *r* was calculated:

$$r = (y/x)^{1/(35-20)} - 1$$

In order to derive the corrected TTO values, *r* was then applied to the number of years provided by the respondents in the TTO such that each future year is multiplied by a function of *r* (Severens and Milne, 2004).

Corrected value = uncorrected value * $(1/(1+r)^n)$, where *n* is the number of years the value is discounted for. Notably, in this context, it is equal to the uncorrected value.

For example, suppose a respondent indicates they would be indifferent at 17 years in full health and 20 years in a given health state (uncorrected TTO value: 17/20 = 0.85) and they have a discount rate of r=0.0015 calculated from their responses to questions 1 and 2.

The respondents corrected value for 17 years (the TTO value ‘numerator’) using their discount rate is calculated as follows:

$$\text{Corrected numerator} = 17 * (1/(1+0.0015)^{17}) = 16.6$$

The same procedure is then carried out with the 20 year (the TTO value ‘denominator’) such that:

$$\text{Corrected denominator} = 20 * (1/1+0.0015)^{20}) = 19.4$$

The corrected TTO value is then:

$$16.6/19.4 = 0.86$$

It was anticipated, based on previous empirical findings, that respondents would be willing to accept being ill for a longer period if the illness can be delayed, assigning less weight to later years (i.e. discounting these years). These respondents are said to have *negative* discount rates whereas respondents who would be willing to be ill for longer at a future period have *positive* discount rates (i.e. less value will be assigned to future years, the more common assumption)[1].

---

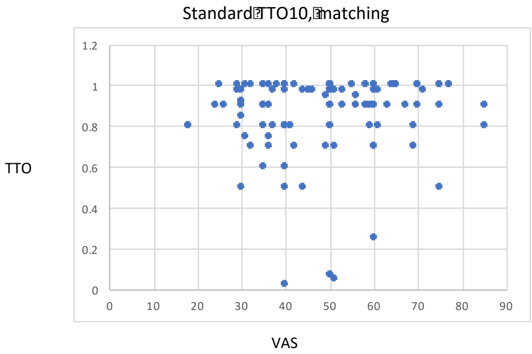
[1] It should be noted that hyperbolic discounting, as opposed to the constant, exponential discounting measured in this study, has frequently been cited as a better approximation of respondent preferences over time (Thaler, 1981; Myerson and Green, 1995). Whereas constant discounting assumes that future outcomes are discounted at a constant rate, hyperbolic discounting implies that gradually less value is allocated to outcomes occurring in the future, and thus discount rates decrease over time (Attema, 2012).

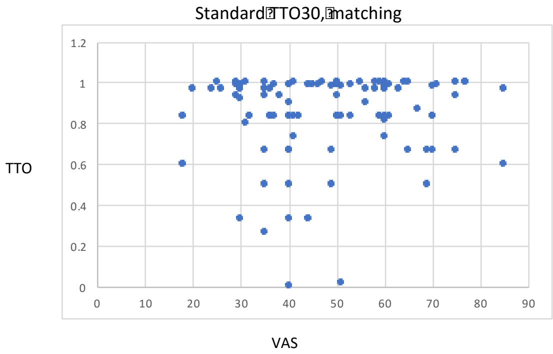
Appendix 2-3: Descriptions of Themes Used in Qualitative Analyses

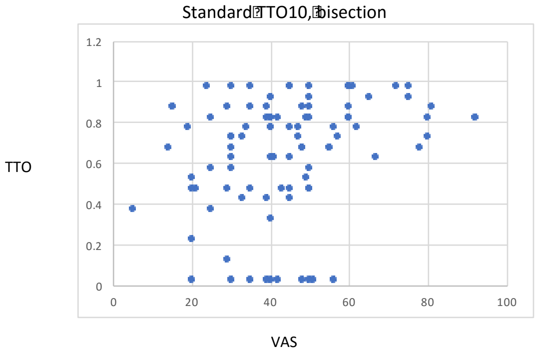
Response Category	Sample Qualitative Evidence
HRQoL vs Longevity	"I was thinking about my quality of life when choosing how many years I would prefer to live in full health compared to living in bad health. I would prefer to die sooner in good health than live longer in bad health. I think living in pain and poor health just isn't worth the extra years." (MS)
	"I arrived at my decision by imagining how my health would affect my daily life and the extent to which it would be debilitating. I also thought about what I would do with an encumbered full lifespan versus a shorter lifespan with unencumbered health. I weighed quantity of life against quality of life." (MA)
Strong preference for length of life, longevity-related goals (e.g. family)	"I was thinking that I wouldn't give up any years for the sake of my wife and kids." (MA)
	"I'd rather live as long as I could with no health issues but I'd also rather live to a certain age even with health issues." (MA)
	"I am 60 my daughter is 8. Getting to her 18th birthday was critical no matter what the health state. Her age was a driver of staying alive longer." (BA)
Health effects on self (health goals, e.g. activities, enjoyment)	"I was thinking about the fact that life doesn't necessarily become less valuable just because it is more difficult. At this time of my life, it seems to me that more time is almost always better. I have known many people who had serious chronic illnesses or were severely disabled who were still able to receive pleasure and a sense of purpose from life, and also to give joy to those around them. I would have to be in a constant state of torturous agony to be willing to part with my life." (MS)
Health state effects on others (e.g. burden)	I was thinking about the burden that I would be for other people. I wouldn't want people to spend their lives taking care of me. I wouldn't want to struggle to take care of myself." (MS)
	"I was thinking about how old my children would be if I chose to die earlier. I was also thinking about how I would not want other people to take care of me. I don't want to be a burden, but I really want to live to see my kids grow up and get married, etc." (MA)
MET	"The longer you live a suffering life, the unhappier you will be so sometimes it is better to go earlier, without the suffering." (MA)
	"Quality of life is very important to me. I don't want to extend my life if that means living in varying degrees of misery." (BA)

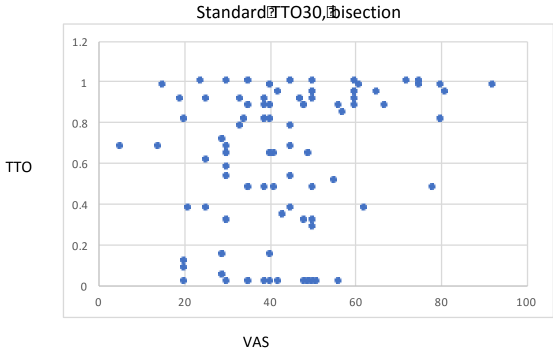
MA matching age-frame respondent, MS matching standard-frame respondent, BA bisection age-frame respondent, BS bisection standard-frame respondent

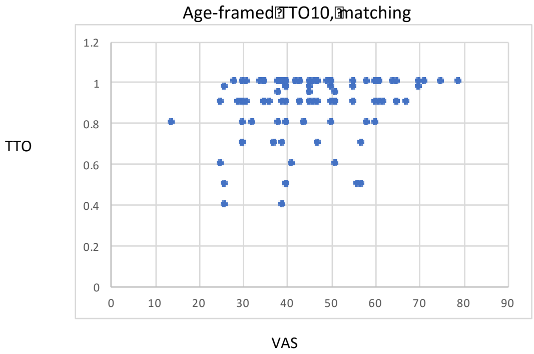
Appendix 2-4: Scatterplots of Individual Respondent Values for VAS vs 10-year TTO (Health State D) and for VAS vs 30-year TTO

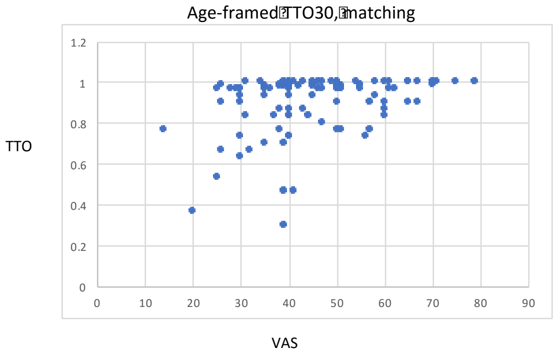


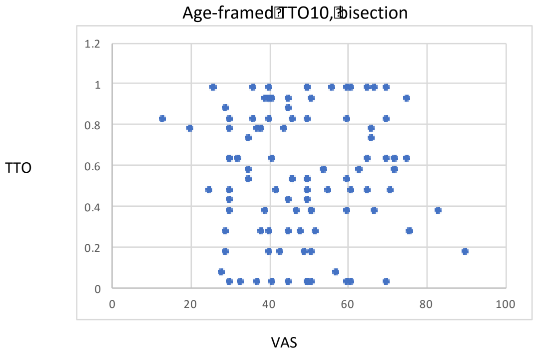


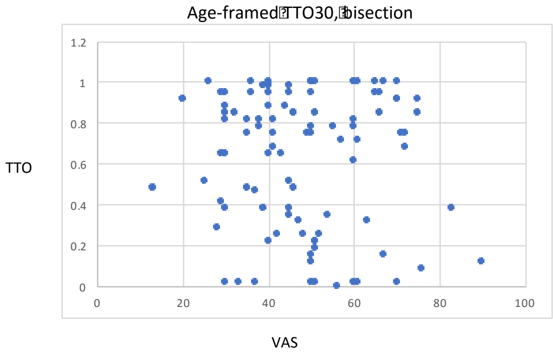






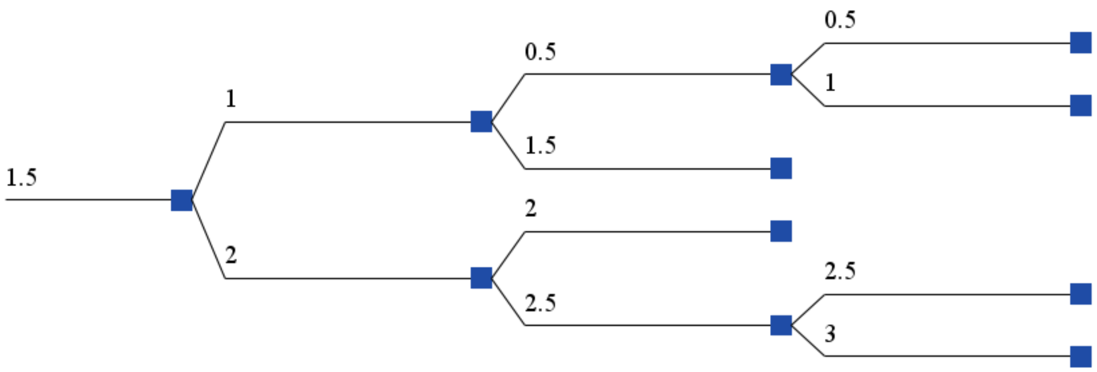




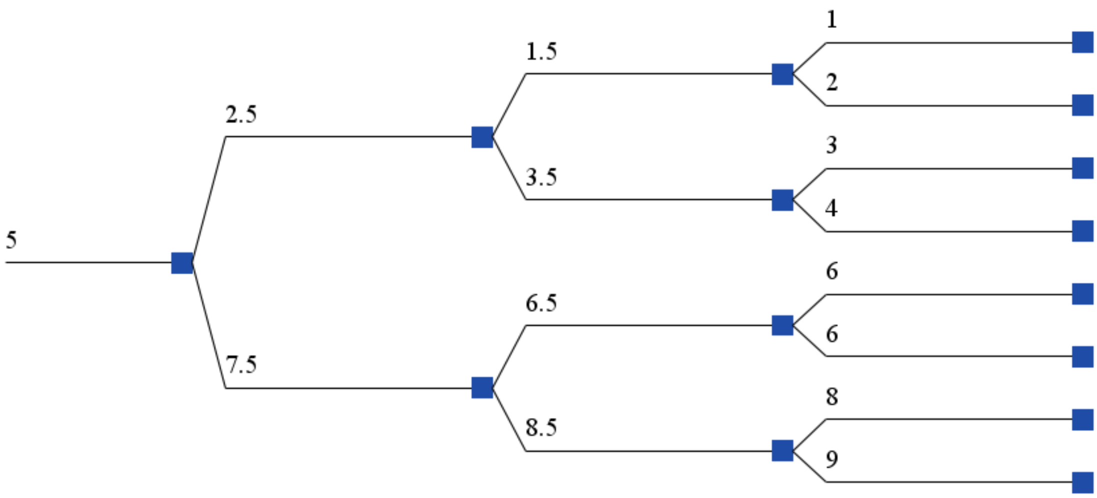


Appendix 4-1: Bisection Iteration Sequences

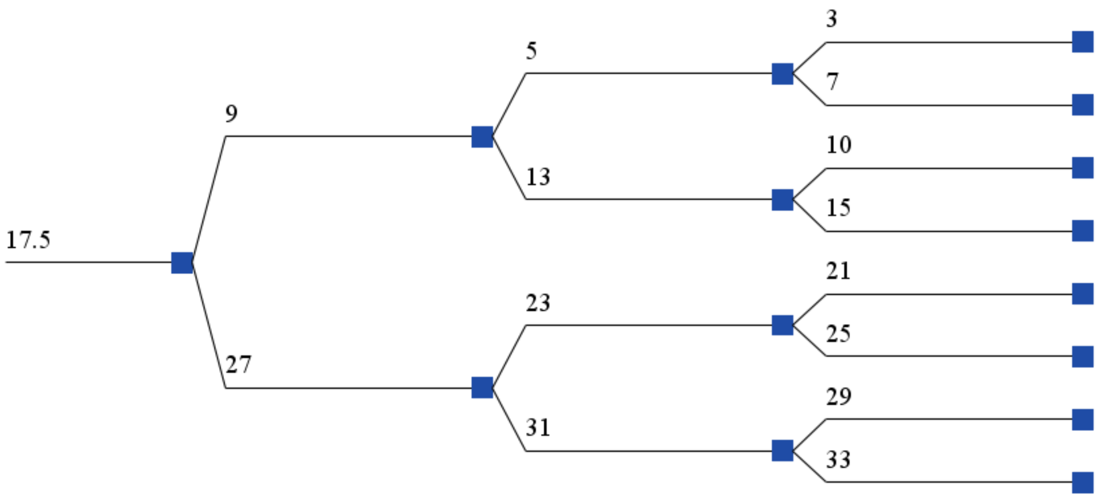
3-year TTO



10-year TTO



35-year TTO



Appendix 4-2: Unadjusted TTO Values

Table A4-1: Unadjusted median TTO values

	Standard TTO			Reverse TTO		
	3-year	10-year	35-year	3-Reverse	10-Reverse	35-Reverse
Unwell						
Matching	0.67	0.8	0.71	0.36	0.57	0.67
Bisection	0.67	0.6	0.43	0.47	0.47	0.48
Back Pain						
Matching	0.67	0.8	0.74	0.41	0.60	0.68
Bisection	0.67	0.7	0.72	0.47	0.67	0.68

Table A4-2: Wilcoxon Signed Ranks Tests between standard and reverse TTO values

	Unadjusted values		
	3	10	35
Unwell			
Matching	<0.001	<0.001	0.04
Bisection	0.02	ns	ns
Back Pain			
Matching	0.005	0.001	ns
Bisection	ns	ns	ns

Note: ‘ns’ indicates the relationship was not significant