TESTING A NEW METHOD FOR ESTIMATING THE MONETARY VALUE OF A QALY

By

CARL TILLING

B.A., University of Sheffield, 2006

M.Sc., University of Sheffield, 2007

A THESIS SUBMITTED IN PARTIAL FULLFILMENT FOR THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

THE FACULTY OF MEDICINE

(School of Health and Related Research)

Table of Contents

Chapter Three – Feasibility, Reliability and Validity of the Contingent Valuation and Time Trade-Off Methods

Chapter Four – Exploring a new method for estimating the monetary value of a QALY: initial testing (survey 1)

Chapter Five – Further Testing of the TTO method of estimating the Monetary value of a QALY: A pilot study to inform a UK based interview study

Chapter Six – Further Testing of the TTO method of estimating the Monetary value of a QALY: A UK based interview study

Chapter Seven – General Discussion and Conclusion

Abstract

Objectives

The objective of this thesis is to develop and test a new method, based on Time Trade-Off (TTO), for the estimation of the monetary value of a QALY (MVQ) informed by public preferences.

Methods

Two new questions are developed to estimate an MVQ which ask respondents to trade off length of life to either increase their income, or avoid a decrease in their income. These questions are initially tested through a Dutch online survey with 321 members of the Dutch general public. The questions are further tested through a small scale pilot study, followed by a UK based interview study with 100 members of the general public. In addition, two further questions are also developed and tested in the UK study, which are more closely aligned with the concepts of Willingness to Pay and Willingness to Accept.

Results

In the Dutch online survey there were a large number of respondents who were not prepared to trade any time to increase their income (or avoid a decrease). Furthermore, some respondents traded too much time, which led to negative MVQ estimates. The prevalence of these responses reduced in the UK based interview study but they were still problematic. Despite this, the questions did appear to be feasible for respondents to complete and were sensitive to scale, particularly in the UK study.

Conclusion

The evidence tentatively suggests that at least some of the respondents stating an infinite preference for length of life over income, were giving a true statement of preference. The questions could potentially be improved by either decreasing the total value of what is being given up, or by increasing the total value of what is being gained. This could potentially be achieved by extending the time horizon in the exercise.

List of Tables

List of Figures

List of Appendices

Acknowledgements

I would not have been able to produce this thesis without invaluable support from many sources. Most importantly I would like to thank my supervisors, Aki Tsuchiya and John Brazier. I wouldn't have embarked upon this challenge without Aki's encouragement and somehow she continued to see potential in me throughout my studies. Aki's eye for detail is unrivalled and through her rigorous enforcement of deadlines I have been able to complete this thesis promptly. John has a wealth of knowledge and experience to draw upon and has continually driven up the standards of my work.

I would like to thank my Dutch colleagues, Werner Brouwer, Marieke Krol and Job van Exel. My stay in Rotterdam was a fruitful and rewarding experience, and their continued collaboration has brought an added perspective to the development of my work. Their scrutiny of my output has been a key factor in enabling me to reach the standards required to make successful journal submissions.

I am grateful to Allan Wailoo and Phil Shackley for providing feedback on versions of my work that were submitted to the HEDS discussion paper series. Jenny Roberts in the Department of Economics was always on hand to give econometric advice, which I could not have managed without. I would also like to thank Lisa Gold and Paul McNamee who discussed versions of my work at HESG meetings. I am appreciative of comments received from anonymous referees at *Value in Health*. Simon Dixon and Helen McTaggart-Cowan were kind enough to lend me copies of their PhD theses.

The UK data collection was completed by Oxford Outcomes. I am indebted to Cicely Kerr whose organisation enabled the project to run smoothly. I also thank the interviewers who applied themselves to a rather unfamiliar and unusual task with enthusiasm. The visual aids used by the interviewers were brilliantly created by the print unit at the University of Sheffield who relished the opportunity to indulge their creative side. Most importantly, I would like to thank the respondents without whom the research would not have been possible.

I was lucky enough to receive an ESRC studentship for the full three years of my study. Finally, I must thank family and friends. Catherine Bunnell's creative skills were invaluable in helping me construct an ad-hoc visual aid for use in the pilot study. Being surrounded by people who understand the perils of a PhD is a great comfort so I must thank the following PhD students: Jenny Willson, Kostas Lagos, Penny Watson, Aurora Ortiz, Heather Brown and Rob Hodgson. I am forever indebted to my parents who put a great emphasis on my education from an early age and who always support and encourage me in whatever I embark upon. Finally, I must thank Thomas Butler for vital support.

CHAPTER ONE

Introduction:
The methodology of economic evaluation and the cost effectiveness threshold

1.1 Introduction

Within the U.K. the role of health economics in health care allocation decisions has grown rapidly since the creation of the National Institute for Health and Clinical Excellence (NICE, 2008), an independent health technology assessment organisation, in 1999. This organisation provides guidance in three areas: the use of health technologies in the NHS, clinical practice, and public health (www.nice.org.uk). Similar bodies provide guidance on the use of health technologies in other countries, for example the Dutch Board of Insurers in the Netherlands (CVZ, 2006). In order to provide guidance on the use of health technologies the methods of technology assessment have grown ever more complex through ongoing research and development.

Economic evaluation is the tool used by health economists to inform the allocation of scarce resources. From a welfarist perspective Cost Benefit Analysis (CBA) is the theoretically ideal approach, and it includes all relevant costs and benefits in monetary terms. Under this approach if the benefits are greater than the costs the intervention should be provided. However, partly due to ethical concerns surrounding attaching a monetary value to human life (Mooney, 1980), Cost Effectiveness Analysis (CEA) has become the more widely used method of economic evaluation in the health care sector. CEA typically uses natural units of measurement, such as cancers detected. However, in order to facilitate comparison between treatments with different outcomes, a generic outcome measure was developed, the Quality Adjusted Life Year (QALY; see 1.2.2).

While in CBA there is a clear decision rule, that if the benefits are greater than the costs the intervention should be provided, in CEA there is no clear decision rule. CEAs simply produce a cost per QALY value, and the decision maker must then decide how much to pay for one QALY. Therefore, it has been argued that a Monetary Value of a QALY (MVQ), informed by general population preferences, would be useful to calculate the benefits in monetary terms and interpret the results of a CEA within a welfare economic framework (Phelps and Mushlin, 1991).

This chapter provides the background and context for the thesis. The following section outlines the welfare theoretic roots of CBA, and the methods of attaching a monetary value to health benefits in this method are discussed. Section 1.2.2 outlines the method of CEA with the QALY as the benefit measure. Section 1.3 outlines and compares the methods for eliciting a value for the quality adjustment in the QALY, since these methods form the basis of an exploratory method developed later in the thesis. The debate surrounding appropriate cost effectiveness thresholds is reviewed in section 1.4, since this is the key issue addressed in this thesis. The objectives and structure of this thesis are then outlined in sections 1.5 and 1.6 respectively.

1.2 The methods of economic evaluation

The analysis of the allocation of scarce resources is the key objective of economics, and is perhaps more important in health economics than any other discipline since any allocative inefficiency is likely to reveal itself in the form of lost lives and/or reduced quality of life. Therefore new interventions must be appraised in a rigorous manner. Furthermore, prior to the establishment of NICE in 1997, allocation decisions were made at the local level and decisions appeared to be made in a rather arbitrary fashion causing what was known as a 'post-code lottery' (Court, 1995). Therefore, decision makers (such as NICE in the UK, and CVZ in the Netherlands) now require appraisals to be made in a transparent and systematic fashion. This is the role of economic evaluation. The two main methods of economic evaluation are CBA and CEA.

1.2.1 Cost Benefit Analysis

The health care industry initially adopted (Klarman, 1967, 1974; Grosse, 1972) the same approach as the rest of the public sector (Foster and Beasley, 1963): Cost Benefit Analysis

(CBA; Mishan, 1971). This approach requires both the costs and benefits to be calculated in monetary terms. CBA is considered theoretically superior to CEA as it allows the results to be interpreted within a standard welfare economic framework (Mishan, 1988; Pauly, 1995). The core of welfare economics (Broadway and Bruce, 1984; Johansson, 1991) is the Pareto Criterion, which suggests that a policy change is socially desirable if everyone is made better off (weak criterion), or at least some are made better off while no one is made worse off (strong criterion). The problem with these criteria is that it cannot inform a policy change that makes some better off and others worse off, and hence is useless in most real world situations. To overcome this shortfall the compensation principle was suggested by Hicks (1939) and Kaldor (1939). Imagine a project that delivers gains for some individuals and losses for others, and also assume that income can be costlessly redistributed across individuals. The Kaldor criterion states that a project is desirable if, with the project, it is hypothetically possible to redistribute income so that everyone becomes better off than without the project. In other words, gainers should be able to compensate losers, although actual compensation is not required by the compensation criteria. This criterion requires the measurement of compensating variation, defined as 'that sum of money received which, following a welfare change, leaves him at his original level of welfare' (Mishan, 1977).

The Hicks criterion states that the losers should not hypothetically be able to bribe the gainers to prevent the project from going ahead. This criterion requires the measurement of equivalent variation, defined as "the sum received by or from an individual which (if he is denied the change in question) leaves him as well off as if he had the welfare change' (Mishan, 1977). Both compensating and equivalent variation are measures of consumer surplus, the appropriate measure depends on whether a programme is being introduced or removed and whether an individual is a gainer or a loser from that project, as will be discussed below in the context of contingent valuation (CV).

The main difficulty in CBA is in attaching a monetary value to human life. There are three ways of doing this: human capital, revealed preferences and stated preferences. Under the human capital approach (Weisbrod, 1961) the value of the programme is assessed in terms of the present value of future earnings (discussed further in Chapter 2). The revealed preference approach observes individuals' real life wage-risk trade-off to infer a value of life

(Marin and Psacharopoulos, 1982). Finally, stated preference studies (most commonly contingent valuation, CV) ask members of the general public their Willingness to Pay (WTP) for a given treatment or the alleviation of a health condition, or their Willingness to Accept (WTA) to forgo a given treatment (see Olsen and Donaldson, 1998). A compensating variation approach to valuing a project seeks to elicit money transfers that return individuals to their original level of utility following the change in question. In this approach, if a project is being introduced the WTP of a gainer for that project to go ahead produces their compensating variation for that gain i.e. the loss of money that returns them to their original level of utility. The WTA of a loser, if the project proceeds, produces their compensating variation for that loss i.e. the monetary gain that returns them to their original level of utility. If the WTP of the gainers is greater than the WTA of the losers then the project should proceed.

Alternatively, the project in question could be valued through equivalent variation, which seeks to elicit money transfers that move individuals from their original level of utility to the new level of utility if the project were to go ahead. In this approach losers are asked their WTP to avoid the change happening and move them to the level of utility they would have achieved had the change gone ahead. Gainers are asked their WTA to forgo the change and move them to the level of utility they would have achieved had the change gone ahead. In this approach if the WTA of the gainers is greater than the WTP of the losers the project should proceed. If a program is to be removed the appropriate questions to elicit compensating and equivalent variation for gainers and losers changes (see O'Brien and Gafni, 1996). In theory, when income effects are small, the values elicited through WTP and WTA should be equivalent (Willig, 1976). However, empirical evidence has shown that WTA values often exceed WTP values (Fernandez *et al.* 2010, see section 3.2.2.5 for further details).

As well as the distinction between WTP and WTA, and whether to use compensating or equivalent variation, there are a number of other design considerations in CV studies. There are a number of possible elicitation formats:

• **Open Ended** – individuals are asked to provide the maximum amount they are willing to pay. This value is both unbounded and unprompted (e.g. Johannesson and Jonsson 1991).
• **Payment Card** – Subjects choose a value from a predetermined and ordered list. All subjects choose from the same list (e.g. Mitchell and Carson 1984). This format is also known as Payment Scale.
• **Random Card Sorting** – A recent alternative to the payment card approach (Carthy *et al.* 1999; Chilton *et al.* 2004; Smith 2006). This method presents individual amounts on separate cards. The cards are then shuffled in the presence of respondents and cards are drawn one at a time in a random order. As each card is presented respondents are asked to sort them into one of three categories: amounts they are sure they would pay, amounts they are sure they would not pay, and amounts they are unsure about.
• **Discrete Choice** – each subject is offered an accept/reject choice at a single, predetermined offer value. Different subjects will receive different offers (Bishop and Heberlein, 1979). A further development of the discrete choice method is the "double bounded" approach which asks respondents a second accept/reject choice (Hanemann, 1985; Carson, Hanemann and Mitchell, 1986). Depending on the response to the first question, the second value presented will be either higher or lower.
• **Bidding Game** – the investigator suggests WTP values that the subject either accepts or rejects and continues to make higher or lower offers depending on whether the subject accepts or rejects the previous offer (e.g. O'Brien and Viramontes 1994).

The relative strengths and weaknesses of these approaches are reviewed in Chapter 3. The second key aspect of study design is the payment vehicle. The payment vehicle refers to the way in which the hypothetical WTP payment is to be made. The most common payment vehicles used include direct out of pocket expenditure, taxation, private insurance premiums and voluntary donations. The payment vehicle adopted must be appropriate for the jurisdiction in which the questions are being asked to make the contingent market as realistic as possible, and avoid potential bias. More details of the effect of the payment vehicle on results can also be found in Chapter 3.

The key advantage of CBA is that as both costs and benefits are measured in monetary terms there is a clear decision rule: if the benefits are greater than the costs the program should be implemented. This is because wealth could potentially be redistributed so that at least one person is better off without anyone being worse off. A further benefit of CBA is that the results from a health care CBA can be compared with any other area of government spending such as road safety or pollution prevention. If the results are displayed in Net Benefit terms (Benefits minus Costs) easy comparison can be made across programmes and sectors.

1.2.2 Cost Effectiveness Analysis

Many decision makers and medical practitioners were uncomfortable attaching a monetary value to human life in the way required by CBA (Weinstein and Fineberg, 1980; Mooney, 1992). This led to a movement away from CBA in favour of CEA. Within CEA benefits were traditionally measured in uni-dimensional natural units. For example Sculpher and Buxton (1993) compared treatments for asthma in terms of cost per episode-free day. The problem with this approach is that the cost-effectiveness of this treatment can only be compared with other treatments that prevent asthma episodes. Therefore, a specific form of CEA was developed using a multi-dimensional benefit measure: the Quality Adjusted Life Year (QALY; see Weinstein and Stason, 1977). Note, this is sometimes referred to as Cost Utility Analysis (Sinclair *et al.* 1981), but the terminology CEA is used throughout this thesis. QALYs are calculated by multiplying the duration in any given health state by the corresponding values for the health-related quality of life in each period. These values are anchored at 1 for 'full health' and 0 for 'being dead or equivalent': values less than zero represent states considered to be worse than dead. The results of a CEA are presented in relation to the next best alternative as an incremental cost-effectiveness ratio (ICER; Black, 1990) which looks as follows:

(Cost of new treatment - Cost of Existing Treatment) / (Benefit of new treatment - Benefit of existing treatment)

In future will simply be referred to as the C/E ratio. The task of decision makers is then to decide how much they are prepared to pay for a gain of one QALY. While CBA addresses allocative efficiency (i.e. what is worth doing), CEA (with the QALY measure) addresses technical efficiency i.e. what is the most cost-effective way of providing a given output (i.e. a QALY). Therefore, in order to decide whether an intervention is worth doing, CEA with QALYs is not enough; one needs a threshold to come from elsewhere

If a WTP per QALY value were elicited from members of the general public the benefits in a CEA could be monetised and the results could be interpreted in a standard welfare economic framework as in a CBA (see 1.4 for more details).

The ways in which the quality adjustment in the QALY can be derived will now be outlined.

1.3 Eliciting a value for the Quality Adjustment in the QALY

There are three main methods for eliciting preferences to yield a value for the quality adjustment in the QALY (Torrance, 1986): Visual Analogue Scale (VAS), Standard Gamble (SG) and Time Trade-Off (TTO). VAS is the simplest of the three approaches and uses a scale like a thermometer ranging from "best imaginable health" (100) to "worst imaginable health" (0). Respondents are asked to rate a given health state by placing it on the scale relative to these two points. In order to scale between 0 and 1, where 0 is dead and 1 is full health, respondents are also asked to position "immediate death" on the scale. The given health state (i) would then receive a VAS score of (x-d)/(100-d), where x is the scale placement of the health state and d is the position of immediate death on the scale. Note, a respondent can have a value <0 if they consider the state to be worse than dead.

Under the SG approach individuals are presented with two alternatives, the first being a certain state and the second being a gamble. For chronic states (i.e. followed by death) preferred to being dead the first option is to live in chronic state i for time t with certainty, and the second option is a gamble which offers full health for time t with probability p or immediate death with probability (1-p). In both alternatives, duration t is followed by death. The value p is varied until the subject is indifferent between the two alternatives and then the preference value for state i is simply given by p. For chronic states considered worse than being dead the first option is to die immediately (with certainty), and the second option is a gamble which offers full health for time t with probability p or chronic state i (worse than being dead) for time t with probability (1-p). Once again the value of p is varied until the subject is indifferent between the two alternatives and the preference value for state i is then given by –p/(1-p).

While the VAS is simple and cheap to administer it entails no sense of sacrifice. The SG is considered by many Health Economists (though not all) to be the benchmark for measuring cardinal health state utilities. It originates from the axioms of expected utility as set out by von Neumann and Morgenstern (1944) and as such has a strong grounding in risk and uncertainty. However, the approach is often considered too complex for respondents to give meaningful responses, and has been found to be internally inconsistent (i.e. respondents gave inconsistent responses when completing a one-stage gamble and a two stage gamble, see Bleichrodt, 2001). The focus of this thesis is on the TTO method, which was developed by Torrance *et al.* (1972) to avoid the use of probabilities. The method consists of a trade off between length and quality of life. Participants in TTO exercises are presented with a choice between two scenarios. In the case of a TTO valuation for a chronic state better than dead one of these scenarios consists of living for a given fixed period of time (t_j) (e.g. 10 years) in the health state to be valued, followed by death. The other scenario (the 'trading scenario') consists of a shorter period of time (t_k) in full health, followed by death. The value of t_k is varied in the trading scenario until a point of indifference between the two scenarios is found and then (assuming zero temporal discounting) the utility value for that health state is given by the value t_k/t_j. For a review of the treatment of states worse than dead in TTO exercises see Tilling *et al.* (2010).

Studies have attempted to directly compare the three different methods. Torrance (1976) found that the TTO method was the easiest to complete and had the highest test-retest reliability (some respondents were re-interviewed at a later date). Dolan *et al.* (1996a) compared SG and TTO and found that TTO performed better on completion rates, test-retest reliability and logical consistency (these terms are outlined further in Chapter 3). Bleichrodt and Johannesson (1997) compare the three methods against direct ranking using Spearman's rank correlation coefficient. The coefficient was highest for TTO suggesting that if one considers direct ranking to be a good benchmark, TTO is perhaps the preferable method. The feasibility, reliability and validity of the TTO method, in comparison with the CV method is outlined in Chapter 3.

1.4 The Cost Effectiveness Threshold

The problem with CEA (with the QALY as the benefit measure) relative to CBA is that there is no clear decision rule. The analysis produces a cost per QALY figure but the decision maker is then left to decide how much they are willing to pay for one QALY. Three broad approaches have been proposed for determining the cost-effectiveness threshold: (i) it should be inferred from previous decisions; (ii) it should be set so as to exhaust an exogenously determined budget and maximise the objective function (e.g. health); and (iii) it should be set so as to determine the optimal health care budget (Culyer *et al.* 2007). Inferring the threshold from previous decisions may be problematic as the threshold might change over time and, as acknowledged in the NICE guidelines, previous decisions may have taken into account 'other considerations' (e.g. the population receiving the technology) which are not relevant for the decision under consideration (NICE, 2008).

In the UK, the only agency that explicitly attempts to use a threshold of any kind is NICE. The threshold is designed to reflect the opportunity cost i.e. to help the NHS maximise health given its exogenously determined budget. The government allocates the health care budget and NICE uses a threshold range to help inform its decisions, whereby interventions below £20,000 per QALY are usually funded and those above this also need to be justified in terms of the degree of uncertainty surrounding the estimate, the particular features of the condition and population using the technology, the innovative nature of the technology, and when appropriate, the wider societal costs and benefits (Rawlins and Culyer, 2004). Typically interventions with a cost per QALY greater than £30,000 will not be funded. Theoretically this shadow price would be calculated by estimating a cost per QALY figure for all interventions provided by the NHS. These treatments would be listed from most cost-effective to least cost-effective and one could work down the list until the budget was exhausted, hence identifying the threshold. However, in reality there is no information on the cost per QALY value for many interventions provided by the NHS, making the implicit threshold adopted by NICE seem rather arbitrary. Furthermore, when a treatment falls under this threshold and is approved for implementation NICE gives no instructions to Primary Care Trusts regarding which treatment they should cease to provide in order to free up funds for the new treatment. This means the current system has failed to completely eliminate the 'post-code lottery' mentioned earlier. NICE has acknowledged this weakness and commissioned research to look at local disinvestment decisions (Appleby *et al.* 2009), but imperfect information makes it difficult to draw conclusions regarding the consistency of NICE investment decisions, and local PCT disinvestment decisions.

As outlined by Culyer *et al.* (2007), NICE is neither a 'threshold maker' or a 'threshold taker'. In order for NICE to be a 'threshold maker' it would have to have power over the NHS budget, which is beyond its remit. In order for NICE to be a 'threshold taker' the government would have to impose a fixed threshold, and the NHS budget would need to be flexible to accommodate it. Therefore, NICE is essentially a 'threshold searcher'. NICE operates a number of search strategies to allocate the exogenously fixed NHS budget (McCabe *et al.* 2008). In collaboration with the Department of Health, it engages in horizon scanning to explore technologies that probably lie in the zone of substitution. NICE also relies upon a broad consultation process with all stakeholders to identify technologies for investment. The 'threshold searcher' model ignores the argument put forward by Birch and Gafni (2002) that resource allocation decisions should not only consider ICERs in isolation, but should also consider budget impact. Even if the budget impact of a new intervention is neutral, if a new intervention replaces an existing less cost-effective treatment, this would suggest that the threshold for future treatments will be lower. If the new intervention is more costly than the one it replaces, this will require more disinvestment, and hence more efficient current technologies will have to be replaced (than if the budget impact were neutral). Therefore, if the NHS budget is fixed, then the cost-effectiveness threshold for an intervention with a large budgetary impact should be lower than for an intervention with a small impact. This means that in order for NICE to efficiently allocate the budget allowance for dynamic interactions must be made. Relative rates of growth of the budget and the productivity of health care must also be considered.

As pointed out by Donaldson *et al.* (2011) the current NICE threshold was based on best guesses of experts at the inception of the organisation, and has been subject to criticism since the UK Health of Commons Select Committee review of NICE in 2001-2002 (House of Commons Health Committee, 2002). Pressure has been placed on NICE to raise the threshold, such as in the case of life extending cancer drugs for patients in the terminal phase of cancer (Lakhani, 2008; Richards, 2008). Ultimately, this pressure lead to the relaxation of the threshold for technologies that could extend life by at least 3 months in patients with a life expectancy of less than 24 months (NICE, 2009). There has also been pressure placed on NICE to lower the threshold based on fiscal pressures faced by Primary Care Trusts who feel NICE guidance is not affordable (West, 2009).

The alternative to the shadow price approach is to determine a threshold externally, such as by eliciting a WTP per QALY value from members of the general public. It has been argued that if there were a known cost-effectiveness threshold the QALY benefits could be translated into monetary terms and directly compared with the costs (Pehlps and Mushlin, 1991). Hence, the results of past and future CEAs could be interpreted in a manner consistent with welfare economic theory (Johannesson, 1995; although this relies on a number of assumptions which may not hold, see section 2.3.6). Such an approach would clearly have implications for the budget. For example, if the value elicited from the public was higher than the current NICE threshold the budget would have to increase in order for this figure to be applied. This would have inter-sectoral, and hence political, implications. Current budget allocation decisions, made by parliament, are based on broad assessments of the marginal value of activity of a wide variety of public programmes compared to the rest of the economy. In other words, expenditure in one sector is considered in relation to the opportunity costs in another sector. Whether methods for eliciting an MVQ would fully incorporate such, often political, considerations is dubious. Therefore, empirical evidence from studies attempting to estimate an MVQ is more likely to be used to give an indication to parliament as to whether the current NICE threshold approximates public opinion, rather than being rigorously applied as a 'one size fits all' threshold.

1.4.1 Estimating an MVQ

Attempts have been made to estimate an MVQ in a number of ways. Estimates have been derived from the existing value of a prevented fatality (VPF) figure used in the assessment of road safety interventions (Mason *et al.* 2009). Estimates have also been derived through use of CV to elicit the WTP for quality of life improving, life extending and lifesaving treatments (Mason *et al.* 2008). The method of deriving a VPF and how this can be used to elicit an MVQ are reviewed in the following chapter. The studies applying CV directly are also reviewed in the following chapter. Given concerns over the CV method, and large variability in existing MVQ estimates, this thesis explores a new method based on TTO (see Chapters 4-6).

In summary, CEA, with the QALY as the benefit measure, is now the preferred method of economic evaluation in health care, and is now recommended by bodies such as NICE. However, this approach lacks a clear decision rule, leading decision making bodies to apply seemingly arbitrary thresholds. This has sparked research seeking to derive an MVQ. As will be shown in the next chapter these studies have encountered problems and struggled to produce consistent results. In this context, the following sections outline the aims and structure of the thesis.

1.5 Objectives of the thesis

The aim of the thesis is examine a new method to derive an MVQ. Specifically the objectives of this thesis are as follows:

1. Develop a new method for the derivation of an MVQ, based on TTO.
2. Test and refine this method to determine whether its use is feasible and valid to consider how it compares to Contingent Valuation.
3. Highlight areas of future research that can develop and test this method further.

1.6 Structure of the thesis

The current chapter has outlined the methods of economic evaluation in health care and introduced the issues that will be addressed in this thesis. The following chapter reviews how the Value of a Statistical Life (VOSL) has been estimated in the public sector, the strengths and weakness of the methods that have been applied, the results that have been produced and how these have been applied in practice. This is particularly relevant as one approach to deriving an MVQ is to base it on the existing VOSL. The alternative is to use CV, and studies using both methods are reviewed in the next chapter. This will highlight difficulties that have been encountered in existing studies and the range of results that have been produced.

The thesis proposes a new method for estimating an MVQ based on TTO, so Chapter 3 compares CV and TTO in terms of their feasibility, reliability and validity to consider whether TTO presents itself as a viable alternative for the estimation of an MVQ. This chapter will also identify criteria against which the exploratory method in subsequent chapters can be assessed. Chapter 4 tests two new TTO based questions that ask respondents to trade off length of life for income. This is done through a Dutch online survey. Chapter 5 presents a small scale pilot study used to inform further refinement of the questions for use in a UK based interview study. The main UK study is conducted with 100 members of the general public and the results are presented in Chapter 6. Chapter 7 draws conclusions from the research, highlights contributions to knowledge in the area and identifies weaknesses and areas for further research.

Chapter 2

Estimating the value of life to inform policy decisions

2.1 Introduction

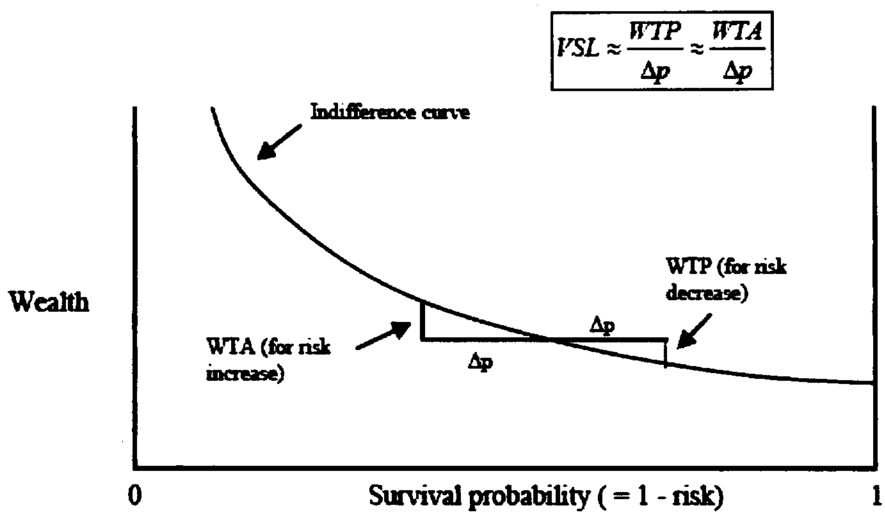
In order to make public sector decisions on risk reducing measures, such as road safety, that align with public preferences it is necessary to estimate the value of life. Furthermore, in health economics interest is growing in whether current QALY thresholds reflect public preferences. The aim of this chapter is to review the ways in which the value of life has been estimated, and how these estimates have been used to inform policy. In public sector economics this has typically consisted of estimating the value of a statistical life (VOSL) through either stated or revealed preferences. In health economics the monetary value of a QALY (MVQ) has typically been estimated either by inferring the value of a life year from the public sector VOSL estimates, or by directly eliciting individual WTP for either quality of life improving, life extending or life saving treatments through the contingent valuation method.

2.2 Value of a Statistical Life

2.2.1 Definition of the Value of a Statistical Life (VOSL)

Essentially the VOSL (sometimes referred to as the value of a prevented fatality, VPF) can be defined as the aggregate WTP (or WTA) across a large group of individuals (ideally those individuals that will be affected by a given public sector intervention) for small reductions in their risk of death which, taken over the whole group, can be expected to prevent one premature death during a forthcoming period. The VOSL is not the 'value of life' in the sense of a sum that any given individual would accept in compensation for the certainty of his or her own death. The VOSL can be considered equivalent to the marginal rate of substitution between wealth and the risk of death. Consider a numerical example: if an individual is willing to pay £100 for a 1/10,000 decrease in his risk of death during the year, his VOSL is £100 divided by the risk of death, so £100 million. The VOSL concept is illustrated in Figure 2.1.

Figure 2.1 – The trade-off between wealth and survival probability (Cropper *et al.* 2011)



In figure 2.1 wealth is plotted on the vertical axis and the probability (p) of surviving a given period is plotted on the horizontal axis. The curved line represents an individual's indifference curve. For each change in survival probability (ΔP), individual WTP or WTA compensation is measured by the vertical distance between the two points on the indifference curve. The VOSL can be calculated as the individual's WTP or WTA divided by ΔP. Note that figure 2.1 assumes that WTP=WTA. In reality whether this holds will depend on the slope of the individuals indifference curve (i.e. the marginal rate of substitution between wealth and income), and their response to gains and losses in light of prospect theory (Kahnemann and Tversky, 1979). The issues of WTP vs WTA is addressed elsewhere in the thesis (see 2.2.4.3). Also note that as the reduction in the risk of death (ΔP) becomes smaller the VOSL estimate can become increasingly sensitive.

Intuitively, one would expect an individual's WTP (or WTA) to increase as the level of risk reduction increases. For small risk changes this relationship should be nearly proportional, but as the level of risk increases the budget constraint will distort this relationship (although the budget constraint should not influence WTA responses). Furthermore, this relationship may be influenced by other factors such as the slope of the indifference curve, income, wealth, age, life expectancy, and current and potential future mortality risk and health status (Cropper, 2011). Factors influencing the VOSL and the problems this creates will be discussed later in this chapter.

As pointed out by Cameron (2010), the terminology 'value of a statistical life' has sparked outrage in American society. Most recently indignation was sparked by a press article published on 10th July, 2008 entitled "an American life worth less today". This article described a decision by the US Environmental Protection Agency (EPA) to revise downwards its estimate of the VOSL from $7.8 million to $6.9 million (see Viscusi, 2009 for a detailed description and history of the EPAs VOSL numbers). Indignant responses to this article included comments from a senator, a front page article in the Washington Post, letters to editors of newspapers, comments on news media websites, and even features in television news satire (as quoted in Cameron, 2010). Previously there had been outrage when it was revealed that the EPA had explored the idea of using a VOSL estimate for seniors that was 1/3 lower than for other adults. In response to this public outcry Cameron (2010) argues that the VOSL should be rebranded "Willingness to Swap for a Microrisk reduction". However, this terminology has yet to catch on and is unlikely to improve the public's understanding of the concept.

2.2.2 How is the VOSL measured?

2.2.2.1 Traditional Productivity Approaches

One way of measuring the value of a lost life is through the value of lost productivity to society through either the *human capital* or *friction cost* methods. Under the human capital approach (Weisbrod, 1961; Rice and Cooper, 1967) lost production (often related to paid work) as a result of morbidity or mortality is valued by measuring time lost from work and multiplying this with the gross wage of the involved individual. Economic theory suggests that under certain conditions for labour market equilibrium, at the margin, gross wages equal the productive value of individuals, so that this multiplication should yield a good estimate of the value of lost production. The relevant period of time over which costs (or savings) are measured is, unless restricted by the time horizon of the analysis, the total period of time in which a person is (un)able to be productive compared to the alternative scenario.

Under the friction cost method (Koopmanschap and van Ineveld 1992; Koopmanschap and Rutten 1993; Koopmanschap *et al.* 1995) the period in which productivity costs occur is limited to the time it takes to replace a worker. The method argues that, from a societal point of view, there are no production losses in the long run, since the production loss in the ill, disabled or deceased worker is cancelled out by a production gain in the new, formerly unemployed, worker. Estimates of the value of productivity change according to the friction cost method to include some additional costs such as the resource cost associated with recruiting and training replacement workers (e.g. advertising the job vacancy).

While these methods are useful for valuing lost productivity, this forms only a component of the VOSL. The approach suffers from a number of serious drawbacks as a measure of the VOSL (Berger *et al.* 2001):

- It violates basic economic fundamentals regarding individual sovereignty, measuring not how much the individual values time, but rather how others value their time.
- It raises equity considerations surrounding the value given to those not in work such as disabled, unemployed, retired etc.
- Perhaps most crucially it does not capture all other reasons for living other than to work e.g. utility from consumption and leisure.

This approach now receives little attention in the literature and will not be considered any further in this thesis.

2.2.2.2 Revealed Preference Approaches

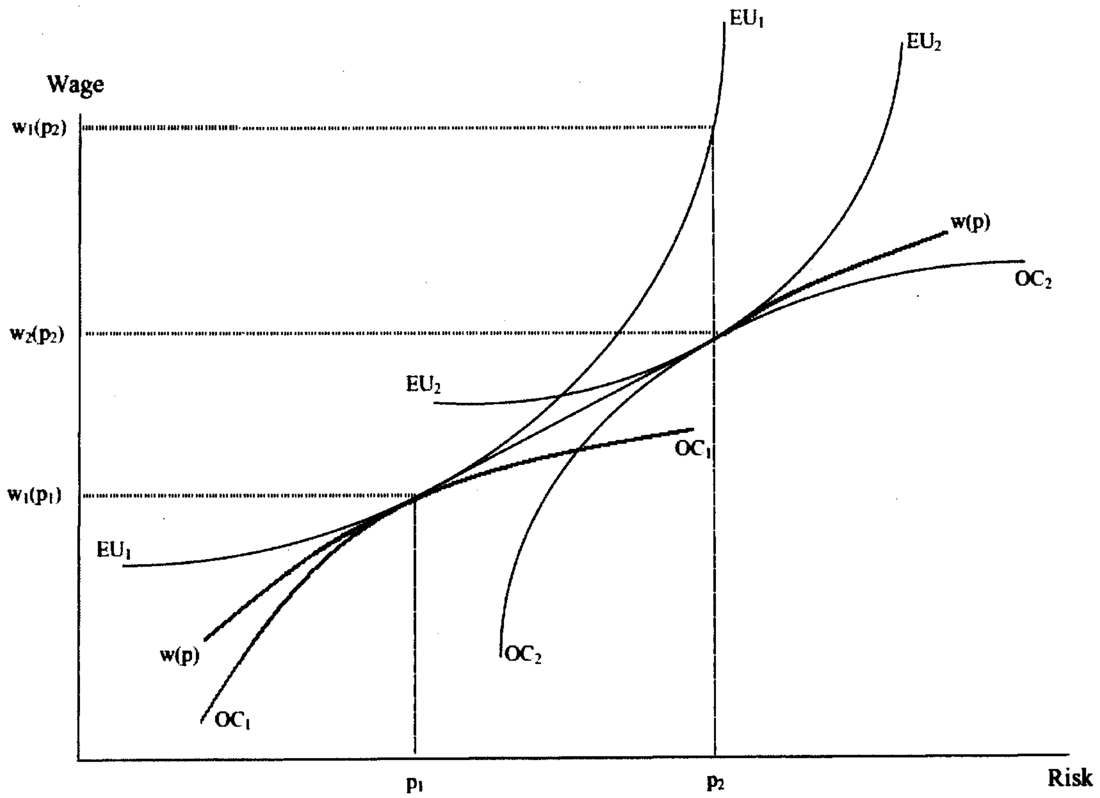
To recap, suppose that an economic agent is willing to accept the consequences of an action that involves an increase in wealth of at least ΔW in return for an increase in the probability of a fatality of ΔP. The $V^*=\Delta W/\Delta P$ is an acceptable trade-off to that agent, and it is the VOSL for the fatality described by the particular fatality risk. The revealed preference approach observes real-world risk taking behaviour to estimate the value of V^*.

Preferences can be revealed through either the hedonic wage method, or the hedonic price method.

The hedonic wage method

The hedonic wage method attempts to quantify compensating wage differentials for on the job risk-exposure in labour markets. The firm's demand for labour decreases with the total cost of employing a worker. The cost of a worker includes not only the wage, but also factors such as training, benefits, and most importantly for the hedonic wage methodology, the cost of providing a safe working environment. Since worker costs increase as the level of safety is improved, firms must pay workers less for a given level of profit following an improvement in safety. Consider the analysis, based on Viscusi and Aldy (2003), presented in figure 2.2. The wage rate if plotted on the vertical axis and occupational risk level is plotted on the horizontal axis. Curves OC_1 and OC_2 represent wage-risk offer curves (isoprofit curves), with wage as an increasing function of risk, for firm 1 and firm 2 respectively. For any given level of risk workers prefer the wage-risk combination from the offer curve with the highest wage level. The outer envelope of these market offer curves is market opportunities locus w(p). The expected utility curves EU_1 and EU_2 show the workers' preferences over risk and wages. The tangency between curve EU_1 and firm 1's offer curve OC_1 represents worker 1's optimal risk choice. Likewise, worker 2 maximises expected utility at the tangency between EU_2 and OC_2. All that are available using real market data are the points of tangency. Hedonic wage models trace at the locus of points of tangency (i.e. curve w(p)) across the labour market.

Figure 2.2 – Market Process for determining compensating differentials (Viscusi and Aldy, 2003)



It should be noted that estimates of the wage-risk trade off curve [w(p)] do not imply how a particular worker must be compensated for a change in his risk level. Referring to figure 2.2 it is clear that if worker 1 was faced with an increase in the risk level he would not move along curve w(p) but instead along their expected utility curve EU_1.

In order to disentangle the wage-risk trade off from other factors that affect wages regression analysis is used. Typically the wage equation takes the following form (Viscusi and Aldy, 2003):

$$w_i = \alpha + H_i\beta_1 + X_i\beta_2 + \gamma_1 p_i + \gamma_2 q_i + \gamma_3 q_i WC_i + p_i H_i \beta_3 + \varepsilon_i \qquad \text{(eq. 2.1)}$$

Where w_i is worker i's wage rate, α is a constant term, H is a vector of personal characteristic variables for worker i, X is vector of job characteristic variables for worker i, p_i is the fatality risk associated with worker i's job, q_i is the non-fatal injury risk associated with worker i's job, WC_i is the workers' compensation benefits payable for a job injury suffered by worker i, and ε_i is the random error reflecting unmeasured factors influencing worker i's wage rate. The terms α, β_1, β_2, β_3, γ_1, γ_2 and γ_3 represent parameters estimates through regression analysis.

The hedonic price method

The hedonic price methodology, also known as consumer market studies, are another form of revealed preferences that examine the observable trade-offs people make between risk and wealth in their everyday consumption decisions (Dionne and Lanoie, 2004). For example, Dardis (1980) uses data on the purchase price of smoke detectors and their effectiveness in reducing the probability of death and injury to estimate the value of statistical life. Atkinson and Halvorsen (1990), as well as Dreyfus and Viscusi (1995), provide estimates based on the price of different safety features on cars and the associated reduction in risk. For more examples, see Dionne and Lanoie (2004) who present a list of 15 consumer market studies performed between 1973 and 1995. Unlike wage-risk studies, consumer market studies have not been repeated many times by different authors, which means that one cannot be as confident in the results produced. Consumer market studies are typically subject to many of the problems encountered in wage-risk studies such as identification of appropriate data, appropriate equation specification and possible omitted variable bias.

2.2.2.3 Stated Preference Approaches

Stated preference studies, which survey respondents about how they would act in hypothetical situations, have become increasingly common. The researcher can specify the level of risk and the scenario can be tailored to specific environmental risks and causes of death. Stated Preference studies take the form of either contingent valuation or choice experiments. Choice experiments consist of either Conjoint Analysis or Discrete Choice Experiments.

Contingent Valuation (CV)

CV studies present respondents (usually members of the public) with hypothetical scenarios about the good (or intervention if in a health care study) under evaluation. They are asked to imagine a contingent market (hence the name contingent valuation) for the good exists and to state the maximum amount of money they would be willing to pay (WTP) to achieve the specified certain outcome. (Alternatively they may be asked the minimum they would be willing to accept, WTA, to go without the programme). Theoretically, contingent valuation method differs from wage-risk studies. Referring back to figure 2.2, while wage-risk studies estimate the wage-risk curve [w(p)], contingent valuation studies a movement along an individual's expected utility curve i.e. how much money would they give up (or accept) in order to maintain their initial level of utility. In other words, revealed preference studies typically examine wage-risk differences between individuals, while CV methods study preferences over risk and income within an individual.

Conjoint Analysis

The survey method of data collection and analysis known as Conjoint Analysis (CA) was developed in mathematical psychology and has a strong theoretical basis (Luce and Tukey, 1964). The method is based on the premises that any good or service can be described by its characteristics (or attributes) and the extent to which an individual values a good or service depends on the levels of these characteristics. Essentially two profiles are presented to the respondent, which typically contain the same characteristics but different levels of these characteristics. Respondents use ranking, rating or discrete choice exercises to represent their preferences for these profiles which then allows the estimation of their respective utility functions. Due to its grounding in random utility theory economists tend to prefer the methodology of Discrete Choice Experiments (DCEs) rather than the ranking and rating methodologies of the conjoint analysis method. If monetary cost is included as an attribute then the ratio of any given attribute to cost shows how much money the individual is willing to pay for a unit change in that attribute (Ryan and Gerard, 2003).

The DCE approach has many of the advantages of CV, including its ability to capture non-health characteristics. One major advantage of DCEs is that whilst CV tells us about the valuation of the whole 'bundle' of characteristics, DCEs help us understand the relative valuations of, or the trade-offs between, various attributes. It is therefore perhaps unsurprising that the use of DCEs has become more popular in the health care field. Examples include examination of individual's preferences for service provision (Scott *et al.*

2003) and for treatment characteristics associated with therapies for osteo-arthritis and prostate cancer (Ratcliffe *et al.* 2004; Sculpher *et al.* 2004). A review of the DCE literature is provided in Ryan and Gerard (2003).

2.2.3 Review of VOSL estimates

2.2.3.1 Search Strategy

As mentioned in the introduction two reviews are conducted, the first considering the estimation of the VOSL and associated theoretical and methodological issues associated with this, and the second considering the estimation of an MVQ. Articles for these two reviews are identified from one initial database search, and further articles are identified through appropriate citation searching and pearl growing.

A review of reviews was chosen as the most appropriate search strategy to capture the necessary methodological and policy debates. This offered an efficient way of capturing the extensive literature in this area.

Three databases were searched: Medline via OvidSP (1948 to week 02/03/2011) using a keyword search of English language review articles; SSCI via Wok (1956 to 19/03/2011) using a topic search of review articles; Econlit via OvidSP (1969 to Feb 2011) using a keyword search (not possible to filter by reviews). These databases were chosen to ensure broad coverage of a range of literatures. The variation in years was determined by the longest period of coverage available in each given database. The following search terms were used:

(QALY OR Life Year OR Year of Life OR Statistical Life) AND (Willingness to Pay OR WTP OR Willingness to Accept OR WTA OR Contingent Valuation OR CV OR Monetary OR Value OR Stated Preference OR Revealed Preference)

This gave a total of 552 results. Titles and abstracts were initially screened to identify articles that were relevant either to the estimation of a VOSL or an MVQ. This process lead to the exclusion of 454 articles which left 98 remaining. For this particular review, focusing on VOSL estimates, the inclusion criteria was any article that reviewed studies directly estimating a VOSL. Primary research estimating a VOSL was therefore excluded, as were theoretical discussions without any review of VOSL estimates. The full texts of the remaining 98 articles were viewed to apply this criteria and nine articles met the criteria: Miller (2000), Bowland and Beghin, (2001), Mrozek and Taylor (2002); De Blaeij *et al.* (2003); Viscusi and Aldy (2003); Blomqvist (2004); Dionne and Lanoie (2004); Kochi *et al.* (2006); Bellavance *et al.* (2009). Two further review were identified, Miller (1990) and Viscusi (1993), but these were not included as they have since been updated by the reviews of Miller (2000) and Viscusi and Aldy (2003) which are included.

Viscusi and Aldy (2003) provide a very thorough review of wage-risk studies and highlight numerous issues in the estimation of a value of a statistical life. Through a prospective pearl growing approach a further, more recent review was identified, Cropper (2011). This paper provides an excellent review of both stated and revealed preference studies and the empirical estimates they generate. From this article two further reviews were identified: Lindhjem *et al.* (2010) and Dekker *et al.* (2011). These reviews had not appeared in the initial search because Lindholm *et al.* (2010) is a report of the Working Party on National Environmental Policies for the OECD, and the review by Dekker *et al.* (2011) was not published at the time of the search (cited as 'forthcoming' by Cropper *et al.* 2011).

2.2.3.2 Summary of VOSL estimates

The aim of this section is to outline the range of estimates presented in the review articles identified and the sources of variance that they encounter. The following section will cover specific issues relating to the different methods that the reviews discuss.

Table 2.1 summarises the results of the reviews. The reviews typically employ a meta-analysis approach. Meta-analysis is commonly applied in the health and medical sciences literatures (see Mann, 1994) and involves pooling raw data from a variety of clinical studies to evaluate the relationships between a health outcome of interest and key variables assumed to affect that outcome (Mrozek and Taylor, 2002). A primary benefit of these applications is the increased evidentiary weight of the larger dataset, which incorporates a larger design space than any one study could provide. Regression analysis is employed to control for differences between the studies included to predict a more accurate VOSL estimate than can be achieved by simply observing the results of previous studies. This approach also allows the researcher to study the relative importance of various determinants of variance in the estimates. For example, by including union membership in the regression analysis as an explanatory variable the researcher can study how this affects VOSL estimates (in hedonic wage studies). Some of the studies do take a more statistically simplistic approach to the synthesis of VOSL estimates. Dionne and Lanoie (2004) and Blomqvist (2004) simple take means and medians of the values they identify.

Viscusi and Aldy (2003) and de Blaeij *et al.* (2003) differ slightly from the other reviews as they do not use statistical analysis to estimate a VOSL figure, they only use meta-analysis to consider determinants of variance (focusing on age and income, addressed in the next section). Therefore, the values presented in the Table 2.1 for these two studies represent the range of estimates (or median) produced by the studies included, rather than a value predicted from statistical analysis. Dekker *et al* (2011) also do not estimate a single VOSL figure, but instead consider the effect of risk size and context on VOSL estimates from contingent valuation studies. The key finding is that respondents are WTP more for risk reductions in the context of air pollution than they are in the context of road safety.

Some of the reviews use VOSL estimates from one country (or multiple countries) to produce a VOSL estimate in another country (e.g. Bowland and Beghin, 2001). They do this by including key differences between the countries as variables in the regression analysis (e.g. GPD per capita). Other reviews include studies from multiple countries and produce only one VOSL figure, or present results for national sub-samples.

Further studying Table 2.1, the reviews cover similar time periods and there is clearly some overlap between them. Of the reviews that considered multiple methods, Miller (2000) found that the CV and hedonic wage methods yielded similar results, but both of these methods produced higher results than the hedonic price method. They state that the identification of appropriate wage and risk data is a key issue in hedonic wage studies, while hedonic price studies must interpret behaviour based on perceived risk levels rather than actual risk levels. De Blaeij *et al.* (2003) find that CV studies produce higher estimates than revealed preference studies. The only explanation offered is the hypothetical nature of stated preference studies, but no strong justification for this is given. Presumably they are arguing that what people say they will do and what they actually do are quite different things. In contrast to this finding Kochi *et al.* (2006) find that the hedonic wage method produces larger estimates than the CV approach. How the two reviews produce such different results is unclear. A full list of included articles is not included in Kochi *et al.* (2006) making it impossible to know the level of overlap between the two reviews. The statistical approaches employed do differ, but it is not possible to determine how this is likely to have affected results. Kochi *et al.* (2006) suggest the differences between the method may arise due to different populations and different types of risk being used. The two methods are also subject to biases (see 2.2.4).

Viscusi and Aldy (2003) confirm the finding of Miller *et al.* (2000) that the hedonic price method produces VOSL estimates that are lower than the hedonic wage method. They offer a number of explanations for this finding. Firstly, while the labour market offers a fairly continuous array of wage-risk opportunities the product market typically only offers discrete choices e.g. whether to buy a cycle helmet. In this case the consumers purchase decision only reveals the lower bound of their WTP. Secondly, the hedonic price method may induce selection based on risk preferences. For example, individuals who engage in risky behaviours, such as cigarette smoking and driving without seatbelts, have lower implicit values for injury then do those who do not engage in such behaviour.

Essentially, the results in table 2.1 demonstrate the lack of consensus around the appropriate method for estimating a VOSL, and particularly around which value should be applied. The following section outlines the effect of age and income on the VOSL, before issues with the different methods are discussed.

Table 2.1: Summary of Reviews of the Value of a Statistical Life

Review	Number of Studies included	Time Period covered	Types of Studies	Range of Results	Calculated for which country
Miller (2000)	68	1974-1999	Hedonic Wage, Hedonic Price and CV	World: $650K EU: $2.73 million UK: $2.75 million (all 1995 US Dollars)	
Bowland and Beghin (2001)	33 data points (number of studies unclear)	Unclear	Hedonic Wage	$519K-$675K (1992 US Dollars)	Chile
Mrozek and Taylor (2002)	33	1974-1997	Hedonic Wage	$1.5 -2.5 million (1998 US dollars)	US
Viscusi and Aldy (2003)	Over 100	1974-2001	Hedonic Wage and Hedonic Price	Median $7million (2000 US Dollars)*	US
De Blaeij et al. (2003)	23	1973-1999	Stated and Revealed Preferences for Road Safety	$147,000-$30.8 million (1996 US Dollars)*	Various Developed Countries
Dionne and Lanoie (2004)	35	1973-2001	Hedonic Price, Hedonic Wage and CV	$1.9-8.3 million (Canadian Dollars)	Canada
Kochi et al. (2006)	40	1974-2002	Hedonic Wage and Contingent Valuation	$5.4 million (2000 US Dollars)	Selection of Developed Countries
Blomqvist (2004)	8	1990-2001	Hedonic Price	$1.7 million-$7.2 million* (2000 US Dollars)	US
Bellavance et al. (2009)	37	1974-2004	Hedonic Wage	$5.86 million (Price Year Unclear)	8 countries but majority of studies from US and Canada

Table 2.1 Continued

Lindhjem *et al.* (2010)	63	1973-2008	Stated Preferences	$2.9 million (2005 US Dollars; median recommended for OECD countries)	Multiple
Dekker *et al.* (2011)	27	1982-2005	Contingent Valuation	$130,000 to $33.6 million (2004 US Dollars)*	Values from 14 countries

* Note: These values are based on estimates from other studies not the output of statistical meta-analysis which is not presented

2.2.3.3 The effect of age and income on the VOSL

The reviews consistently identify age and income as key factors determining variance in the VOSL estimates. This clearly has important implications for the application of a VOSL to a public sector intervention aimed at a given population. The observed positive relationship between income and VOSL estimates (in both stated and revealed preference approaches) is most likely explained by the fact that safety is a normal good (Viscusi 1978). Therefore, the VOSL should increase with per capita income. In policy, estimates of the income elasticity of the VOSL are used to adjust the VOSL over time, but agencies do not typically adjust their estimates for cross sectional differences in population sub-groups (Hammitt and Robinson, 2011). The U.S. Environmental Protection Agency (EPA) generally applies a distribution of values to characterize uncertainty about the relevant income elasticity, with a mode of 0.40 and endpoints at 0.08 and 1.00 based on its review of the literature (US EPA, 1999). It measures the change in income using yearly estimates of real per capita GDP (see US EPA, 2006 for an example).

One might expect the VOSL to decrease with age, since older people have fewer years left to live. Indeed some researchers have shown such a relationship (Rosen, 1988). Stated preference studies that address risks to children suggest that reducing these risks may be valued more highly by members of the general public than reductions for adults (Hammitt and Haninger, 2010). However, Blomquist *et al.* (2011), in a stated preference study of the value of preventing asthma risks, find that the relationship of VOSL to age is complex, declining from age 4 to 30, increasing from 30 to 66, and declining over older ages. Krupnick (2007) conducts a review of 28 stated preference studies and finds that only approximately half of them find that the VOSL estimate statistically significantly reduces with age. An inverted U-shaped age relationship has been found in both revealed preference studies (Viscusi and Aldy, 2003; Aldy and Smyth, 2007) and stated preference studies (Jones-Lee, 1985; although this study fails the external validity test in the review by Krupnick, 2007). Aldy and Viscusi (2007) give the following explanation for this relationship. The increasing part of the inverted-U reflects the low consumption individuals have as young adults, and their increasing consumption with age increases their willingness to pay for risk reduction. Their consumption then flattens out as they start saving, so the increase in WTP for risk reduction slows down and eventually peaks. At this stage, the phenomenon that fewer remaining years of life expectancy explains WTP for risk reduction begins to dominate. Elderly individuals, who consume less with each year, may reduce their willingness to pay even faster with age.

From a policy perspective there has been much debate over whether there should be a 'senior discount'. In 2000, Canada employed a VOSL for the over-65 population that is 25% lower than the VOSL for the under 65 population (Hara and Associates, 2000). In 2001, the European Commission recommended that member countries use a VOSL that declines with age (European Commission, 2001). As mentioned earlier the U.S. EPA experimented with the use of a senior discount in 2002 when evaluating the 'clear skies initiative' but was forced to abandon such an approach due to public outcry. In the UK context an ad-hoc group working for the Department of Health to monetise the health effects of air pollution used age weighting. Since air pollution mainly affects over 65s they weight WTP by 0.7 for this group and also adjust for reduced life expectancy (DoH, 1999).

The relationship between age and the VOSL has implications for the estimation of a value of a statistical life year (VOSLY) for use in a healthcare setting. The simplest approach is to use average remaining years of life expectancy nationally. For example, if the VOSL is £1million, and the average remaining years of life expectancy is 40, then the VOSLY is £25,000. In some cases discounting is applied. However, if the relationship between the VOSL and age is not linear then this approach is no longer valid. This is addressed further in the later section on estimating an MVQ (section 2.3).

2.2.4 Issues with the methods

2.2.4.1 Hedonic Wage Method

Risk Data

One of the key difficulties with the hedonic wage method is in obtaining accurate estimates of the risk of death on the job. The standard approach is to use industry-specific or occupation-specific risk measures reflecting an average of at least several years of observations for fatalities, which tend to be relatively rare events. Early U.S. labour market papers used a job related risk measure based on data collected by the Society of Actuaries for 1967 (Thaler and Rosen, 1975; Brown, 1980; Leigh, 1981; Arnold and Nichols, 1983). This dataset provides fatality risk data across 37 occupations. This approach assumes that worker's perceptions of risk are aligned with these statistical measures of risk. If workers perceive the level of risk to be lower than it actually is then inferred WTP will be an underestimate and visa versa. Very few studies have compiled worker's subjective preferences regarding risks. Hamermesh (1978), and Viscusi (1979, 1980) estimated the hedonic wage equation with a dichotomous measure of injury risk based on a worker's perception of whether his or her job is "dangerous". These studies estimated statistically significant coefficients on this variable. However, Mrozek and Taylor (2002) have shown through meta-analysis that studies accounting for workers' perceived risk produce similar VOSL estimates to studies using traditional actuarial datasets. A further problem with use of actuarial data is that it does not distinguish fatalities caused by the job but rather reflects rates of people within a particular job category. Viscusi and Aldy (2003) illustrate this point using the example of actors, one of the highest risk occupations based on actuarial ratings, but who typically face few risks in their profession.

Recent studies (Viscusi, 2004; Aldy and Viscusi 2008; Kniesner *et al.* 2010) have made significant advances in the measurement of job risk by using the BLS Census of Fatal Occupational Injuries (CFOI) and distinguishing risks by occupation and industry. These studies generally use risks for 720 occupation-industry cells (10 occupations and 72 two-digit industries) based on three-year averages of deaths. Several UK studies use data provided by the Office of Population Censuses and Surveys (Sandy and Elliott, 1996; Arabsheibani and Marin, 2000), while others used unpublished data from the U.K. Health and Safety Executive (Siebert and Wei, 1994). However, the potential issue of divergence between workers perceptions and the statistical data still remains.

Wage Data

Clearly, in order to estimate the compensating wage differential an estimate of the wage rate of different industries and occupations is required, which can be combined with the above risk data. This data can be directly collected from workers via surveys (Lanoie, Pedro and LaTour (1995) but typically it is easier to collect this information from existing datasets. Datasets in the U.S. include the University of Michigan's Survey of Working Conditions (SWC), the Quality of Employment Survey (QES), the Bureau of Labour Statistics Current Population Survey (CPS), the Panel Study of Income Dynamics (PSID), and decennial census data. In the UK the General Household Survey has typically been used (Arabsheibani and Marin, 2000). The dependent variable in the majority of labour market analyses has been a measure of the hourly wage. In some cases this will have to be constructed from weekly or annual labour earnings data.

Ommitted Variables Bias and Econometric Isssues

Failure to control for either worker or job characteristics that are correlated with job risk will render estimates of the coefficient on fatality risk (γ_1 in equation 2.1 earlier) biased and inconsistent (Cropper *et al* 2011). If higher-risk jobs have undesirable characteristics not measured by the researcher, the risk variable will tend to capture these characteristics, biasing the coefficient on fatality risk upward. If worker productivity is measured imperfectly and more productive workers accept safer jobs, this will bias the coefficient downward (Hwang *et al.* 1992).

A tricky issue is the inclusion of non-fatal injury. The high correlation between fatal and non-fatal injury can make joint estimation difficult. Some studies attempting to estimate equations with both types of risk in them have found non-significant coefficients on at least one of the measures (Dillingham and Smith, 1984; Kniesner and Leeth, 1991). However, Cousineau, Laccriox and Girard (1992) found that omitting injury risk may cause a positive bias in the estimation of the coefficient on mortality risk, because it captures the additional risk (i.e. omitted variable bias).

Many other authors have studied the sensitivity of estimates of the price of risk to equation specification, including Black *et al.* (2003), Hinternmann *et al.* (2010), Leigh (1995), and Mrozek and Taylor (2002). Correlations between fatal job risk and industry or occupation make estimates of the price of risk sensitive to their inclusion. Leigh (1995) argued that since the coefficient on risk often becomes insignificant when industry dummies are included it is likely that estimated price of risk actually captures inter-industry wage differentials. In their meta-analysis, Mrozek and Taylor (2002) find that studies that include industry dummies obtain significantly lower estimates of the price of risk than studies that exclude them. This is an issue of colinearity that can only really be overcome by collecting better data.

Endogeneity of Risk

Risky jobs are likely to attract a specific type of personality. Statistically this means there exists a correlation between risk and unobserved worker characteristics. This can potentially be controlled for using instrumental variables but finding a good instrumental variable for job risk has proved difficult. Kneisner *et al.* (2010) and Hintermann *et al.* (2010) use past risk levels as instruments for the change in worker risk in their studies. An alternative if panel data is available is to use first differencing or worker fixed effects which will eliminate worker characteristics that change slowly over time from the error term (Kniesner *et al.* 2012; Hintermann *et al.* 2010).

2.2.4.2 Hedonic Price Method

Perhaps the greatest difficulty with the hedonic price method is the endogeneity of risks. Ashenfelter (2005) illustrates this problem using an example of driving speeds. Individuals take risks by driving faster to reduce travel time, which is a form of travel cost. We might consider the relation of speed to fatalities across roads as a measure of the causal effect of speed (or the time it saves) on fatalities. However, both speed and fatalities are both influenced by numerous external factors, such as road condition and congestion level, meaning that the link between speed and fatalities is weak. Furthermore, are drivers fully aware of the risks they are taking when they increase their speed? Add to this the difficulty in obtaining appropriate data and one can perhaps see why wage-risk studies are more prevalent.

2.2.4.3 Contingent Valuation

General concerns surrounding the CV method, such as the vulnerability of results to aspects of study design, are discussed in detail in the following chapter. This section focuses on issues particularly relevant to the estimation of a VOSL.

Perhaps the greatest concern is the alleged insensitivity to scale of the CV method. The qualitative expectation is that WTP should be positively associated with magnitude of risk reduction – assuming risk reduction is a desired good. Furthermore, economic theory suggests that, for reductions in small probabilities of death, illness or impairment, WTP should be nearly proportional to the change in probability (Jones-Lee, 1974; Weinstein *et al.*, 1980). Thus if a reduction in annual mortality from 20 in 100,000 to 18 in 100,000 is valued at $20, then a larger reduction from 20 to 16 in 100,000 should be valued at about $40 (ignoring a tiny effect of diminishing marginal utility of income). However, most CV studies have estimated WTP values that vary less than proportionately to the risk reduction, and so the derived VOSL estimate depends on the (usually arbitrary) choice of risk reduction (Hammitt and Graham, 1999). For example, Jones-Lee, Hammerton and Philips (1985) estimated mean WTP to reduce the risk of dying in a travel accident by 4/100,000 as £137, yielding a VOSL of £3.4 million. Mean WTP for a 75% larger risk reduction (7/100,000) was only 15% greater (£155), yielding a smaller estimated VOSL of £2.2 million.

A possible explanation for this apparent insensitivity to the magnitude of the risk reduction is that respondents cannot adequately comprehend small risk changes. Risk changes are typically kept small in order to avoid the influence of both the budget constraint and the level of baseline risk. Clearly as the risk reduction becomes larger a respondents WTP is going to become influenced by his ability to pay. Economic theory also suggests that the baseline level of risk will influence WTP (Pratt and Zeckhauser, 1996). Hammitt and Graham illustrate this through a game of Russian Roulette. The standard analysis concludes that it makes sense to pay more to reduce the number of bullets in a six-chamber revolver from 5 to 4 than from 2 to 1. However, although the movement from 2 to 1 may have an illusion of superiority because of the relative percentage difference, 50% vs 20% reduction in risk, this temptation is eliminated by a focus on the absolute reduction in the two cases (1/6 = 1/6). If the baseline risk is sufficiently small the effect on WTP should be negligible (Hammitt and Graham, 1999).

As mentioned, problems arise with small risk changes if respondents cannot adequately comprehend them. Baron (1997) found that subjects' willingness to pay for government medical insurance that could cure a given number of people from diseases varied depending on the number of people who could *not* be cured. In a second experiment, WTP for risk reduction was unaffected by whether the risk was described in terms of percentage or number of lives saved, even though subjects knew that the risks in question differed in prevalence. Baron argues this is explained by a general tendency to confuse proportions and differences. Fetherstonhaugh *et al.* (1997) also found that an intervention saving a fixed number of lives was judged significantly more beneficial when fewer lives were at risk overall.

A number of risk communication tools have been developed to assist respondents in comprehending the magnitude of risk reductions, which seem to have had some success. Krupnick *et al.* (2002) show respondents their baseline risk of death (i.e. risk of death in the absence of the program) over the next 10 years by darkening squares on a grid of 1,000 squares. The risk reduction that the respondent is asked to buy is communicated by changing red squares to blue. In this study responses pass the external scope test: mean WTP of respondents faced with a large risk reduction exceeds that of respondents faced with a smaller risk reduction. Corso *et al.* (2001) allow for direct testing of the effect of visual aids to communicate risk by using four subsamples. Three subsamples receive different versions of the visual aid, while the fourth subsample receives no visual aid. The visual aids were variations of a risk ladder on which each rung represents a progressively higher risk level. The results showed that WTP was sensitive to the magnitude of the risk change for the three subsamples with visual aids, but not for the subsample with no visual aid. For one of the visual aids WTP increased proportionately with the size of the risk reduction.

Although risk level is usually very low to avoid it influencing the WTP value, some evidence suggests that the initial risk level still has a significant impact. In their meta-analysis de Blaeij *et al.* (2003) include variables for both the initial risk level and the change in risk in their regression analysis. The coefficients were significant and suggested that WTP was an increasing function of the initial risk level, and a decreasing function of the risk change (due to insensitivity to scale). De Blaeij *et al.* (2003) therefore argue that the attainment of a single VOSL is not feasible from a theoretical perspective.

WTP vs WTA

This issue is addressed in the next chapter so is only touched upon briefly here to outline potential policy implications in the context of the VOSL. Standard economic theory suggests that there should be only very small discrepancies between WTP and WTA (Sugden, 1999). However, there is evidence that the WTA approach produces significantly higher values (see for example Guria *et al.* 2005). When stated preferences have been used to estimate a VOSL WTP has been the dominant method. In their review de Blaeij *et al.* (2003) do identify a small number of studies that have used WTA, and they produce only slightly larger estimates than WTP.

Divergences between WTP and WTA values suggest that members of the public value an intervention differently depending on whether it is being introduced or withdrawn. Guria *et al.* (2005) identify a number of policy scenarios where this may be problematic. Suppose a new regulation aimed at improving safety is evaluated through stated preference WTP based values of safety. The cost benefit ratio is positive and the regulation is introduced. However, sometime later the benefits turn out to be lower than expected. If these actual benefits had been incorporated in the original CBA the regulation would not have been introduced. Nonetheless, rescinding it would deprive people of various safety benefits in return for a saving of resources. The appropriate calculation in this scenario is a WTA based value of safety. If the WTA value is twice the WTP value, and the loss from abolishing the regulation would outweigh any cost savings, a regulation that would not now be introduced remains in place. This suggests that there may be a degree of irreversibility in decisions concerning the introduction of new safety measures.

Another scenario where the WTP vs WTA discrepancy could be problematic is if there is a potential regulation that would decrease the risk to some people but increase the risk for others. For example, consider a scheme to divert traffic from one area to another. Suppose that Area A currently has relatively high traffic density and hence a high road accident rate, while Area B has relatively low traffic density and a low road accident rate. Suppose there is a proposal that will divert traffic from Area A to Area B, thereby reducing fatalities in Area A by 50 but increasing fatalities in Area B by 25 – a net benefit of 25 prevented fatalities. If the WTP based VOSL were $4 million the scheme would be worth $100 million. If the costs of the program were, say, $50 million, evaluating the proposal in this way would generate a benefit:cost ratio of 2.0. But if we take the view that the proposal increases the risks to the population of Area B and that we should use a WTA-based value of $10 million per fatality to represent their preferences, then the benefits (50 avoided deaths valued at $200m) come to less then the costs ($50m plus $250m value of the additional 25 deaths in Area B). Therefore, under a conventional cost-benefit analysis and appealing solely to considerations of efficiency, this proposal would be rejected, while an alternative proposal that would cost the same but prevent fewer deaths in total would be implemented.

2.2.4.4 DCE

DCEs (or Conjoint Analysis) are used far less frequently in the estimation of a VOSL than the CV approach. Telser and Zweifel (2007) ask elderly Swiss subjects at risk of hip fracture to value different attributes of a hip protector using DCE. These attributes were: protective effect (i.e. risk reduction), ease of handling, wearing comfort and out of pocket cost. Through a probit regression the results of this exercise allow the calculation of a respondents' WTP for a given reduction in the risk of hip fracture. As a test of convergent validity the authors use this information to calculate a VOSL for comparison with hedonic wage derived VOSLs. Before making their hypothetical decisions in the DCE, respondents were informed about their risk of breaking the femur as well as the resulting risk of death.

Mortality rates given fracture of the femur were used according to age class to estimate the implied relative reduction in mortality due to this particular cause. This information leads to the calculation of a VOSL amounting to $1.9 million (US) for individuals aged 70-75. In order to make it comparable with the hedonic wage literature this value is adjusted to an age of 40 years using the empirically found minimum and maximum values of the differences between ages 40 and 70 (see Jones-Lee *et al.* 1985; Carthy *et al.* 1999). This leads to a mean value of $3.2 million, which is similar to the European values given in the meta-analysis by de Blaeij *et al.* (2000).

The review by Krupnick (2007; which focuses on the relationship between age and VOSL estimates) identifies 35 studies, of which only seven use DCE. The DCE studies typically produce VOSLs that are higher than the CV studies. Two of the seven DCE studies estimate a VOSL through both CV and DCE (Chestnut *et al.* 2003; Strand *et al.* 2004). Only one of these is accessible (Strand *et al.* 2004). In this study the DCE exercises produced higher VOSL estimates than the CV approach. However, the most important finding is that VOSL estimates were not sensitive to the size of the risk reduction (i.e. WTP varied proportionately with the size of the risk reduction) while this was not the case for the CV questions. The author argues this is encouraging for the DCE method.

Concerns have been expressed over the inclusion of a cost variable (Ratcliffe, 2000; Skjoldberg and Gyrd-Hansen, 2003), which is obviously essential for the estimation of a VOSL. Both studies argue that the results of DCE experiments are likely to be influenced by the choice of levels of the cost parameter. This is essentially a form of range bias as observed in contingent valuation studies, discussed in the next chapter. The number of cost levels is typically limited by a desire to avoid over complicating the study design. Limiting the number of cost levels in this way is likely to reduce the sensitivity of the WTP values across respondents (although this was not observed by Strand, 2004, above – the number of cost levels is not clear). Skoldberg and Gyrd-Hansen (2003), using a large dataset, find that as the range of cost levels applied increases the WTP value also increases.

2.2.5 How have VOSL estimates been used to inform public policy

Viscusi and Aldy (2003) give a good summary of the use of VOSL estimates throughout the world. Regulatory agencies in the US, Canada and the UK have been most prominent in their use of VOSL estimates. In the US the human capital approach was traditionally the favoured method. The impetus for the adoption of the VOSL method was that the values boosted the assessed benefits by roughly an order of magnitude, improving the attractiveness of agencies' regulatory efforts. The Office of Management and Budget (OMB) has responsibility for overseeing and co-ordinating the review of regulatory impact analyses. The OMB has published guidelines for all Federal agencies recommending the use of a VOSL to monetize the benefits associated with rules that change the population's mortality risk (e.g. U.S. OMB, 1996). The US EPA recommended an EPA of $6.2 million (2000 US$), while the US Federal Aviation Administration recommended a value of $3 million in its 2002 economic analyses. This shows the considerable variation in values applied across the public sector.

The three main areas in which the VOSL has been applied are in road safety, the environment and medical technologies. In road safety the VOSL can inform both optimal decisions between possible projects, and also the appropriateness of the budget. For example, imagine there are three possible safety improving projects. The cost of saving a life in these projects is £4 million, £2 million and £1 million. Clearly, the third project is preferable. If the budget were £3 million the second and third projects would be funded but not the first. Is this optimal? If the VOSL is £4 million then all three projects should be funded so the budget needs to be increased. However, if the VOSL is £1 million only the third project should be funded and the remaining resources should be re-allocated to another area of public spending. The approach in environmental assessments is similar. The idea is to estimate fatality risks associated with a particular environmental hazard and compel abatement to the point where the cost of an additional life saved becomes greater than the VOSL.

For use in assessing health care interventions the VOSL can be translated into a VOSLY. The simplest approach is to use average remaining years of life expectancy nationally. For example, if the VOSL is £1million, and the average remaining years of life expectancy is 40, then the VOSLY is £25,000. In some cases discounting is applied. This will be addressed further in section 2.3.

2.2.6 The UK Value of a Prevented Fatality (VPF)

In the UK, the Cabinet Office has provided guidance for economic analyses for the governments' regulatory and policy-making agencies (HM Treasury, 2004). This guidance does not specify the VOSL to be used, but it does recommend careful consideration of the challenges in applying values estimated in the economic literature to potentially different risk and population contexts of the policy. When used in the context of road safety the VOSL is often referred to as the Value of a Prevented Fatality (VPF). Currently the Department for Transport applies a CV based VPF of £1.54 million in 2009 prices (Department for Transport, 2007). Due to the sensitivity of the VPF estimates to the size of the risk reduction (i.e. the insensitivity to scale of WTP), as outlined earlier this current UK VPF was estimated through a 'chained CV/SG' approach (Carthy *et al.* 1999). In the first stage respondents are asked their WTP for the certainty of a complete cure for a given non-fatal road injury. A total of 14 payment cards are used ranging from £50 to £2 million, and the payment method is out of pocket. The second stage determines the ratio of the health state value for being dead to that for non-fatal injuries using SG. The results from the two stages are then combined to obtain the WTP for a reduction in the risk of death. The authors argue that the majority of respondents displayed sensitivity to injury severity in their WTP responses. However, 11% of the sample had to be excluded because their SG responses implied an infinite ratio in the second stage (i.e. they were unwilling to gamble).

The current UK VPF was recently reviewed in a report commissioned by the Department for Transport in 2011 (DfT 2011). In this report the authors see convincing cause to prefer stated preference methods over any other approach, and conclude that the chained approach is in principle superior to other stated preference techniques that have so far been used to estimate WTP-based fatality risks. They do, however, identify a number of outstanding issues. Two studies are identified that present evidence on the chained approach. The first of these was an EU funded project (OECD, 2010) conducted between

2006 and 2009, investigating the Valuation of Environmentally related Health Impacts (VERHI). The UK team considered whether to use some variant of the CV/SG chained approach or whether to use more conventional direct CV questions. The first part of the study, a small scale laboratory test using standard CV questions, found that the same risk reductions were liable to be valued differently depending on the way they were presented, and sensitivity to the size of the risk reduction deteriorated as the outcomes became more unfamiliar. The second stage allows comparison between the chained and direct approaches. Responses to the direct approach were insensitive to the size of the risk reduction and elicited a larger number of non-trades (i.e. would not pay any money). The chained approach elicited valuations for two different injuries. The VPF based on one of the injuries was twice the VPF based on the other injury. Although there were very few non-trades in the initial CV question in the chained approach, there were indications that the inclusion of the SG exercise increased the complexity of the task and led to possible confusion in the minds of at least some respondents.

In the main VERHI study a few changes were made to the design of the chained approach. To encourage respondents to be realistic about what they could and could not actually afford, a 'budget constraint' exercise was inserted ahead of the WTP questions, to try to establish the kinds of amounts people thought they could afford quite easily, or with some difficulty, or with great difficulty, or just could not contemplate affording under any circumstances. Four illness scenarios differing in degrees of severity are used. Results showed that 28% of the sample gave exactly the same WTP for all four scenarios, and 16% of the sample gave the same SG responses to all four SG questions. There was great diversity and skewness of the distribution of SG responses across the samples which is problematic when computing individual level VPFs . Using an individual approach gave arithmetic mean VPFs in excess of £1,000,000,000. The authors opt to present VPFs derived from chaining the means of each stage.

The other study identified by the DfT report that allows the assessment of the chained approach is the UK Social Value of a QALY (SVQ) project (Baker *et al.* 2010) which is addressed further in the following section on the estimation of an MVQ. In brief, the study gave encouraging results for the sensitivity of CV, but encountered the same problems in chaining the CV and SG responses as encountered in the VERHI study outlined above. Ultimately the authors of the DfT report conclude that the chained CV/SG approach is the preferred method to calculate an MVQ but that further research is needed, particularly into procedural effects and the nature of imprecision in CV and SG responses.

2.3 Estimating the Monetary Value of a QALY (MVQ)

As outlined in the introduction there has been pressure both to increase the current NICE threshold, as exemplified by the case of life-extending drugs which ultimately lead to the new weighting of the threshold for end of life drugs (Lahkani, 2008); and to decrease the threshold in light of fiscal constraints (West, 2009). These debates have sparked an increase in research into whether or not the current threshold is aligned with public preferences. The aim of this section is to review existing studies that have attempted to estimate the monetary value of a QALY.

2.3.1 Search Strategy

Articles were identified from the initial search outlined in section 2.2.3.1. Inclusion criteria for this review was any review article that considered the estimation of an MVQ. That could be either a review of empirical estimates or a discussion of theoretical/methodological literature relating to the estimation of an MVQ. The initial database search identified only three reviews that were relevant to this topic (Hirth *et al.* 2000; Richardson and Smith, 2004; Gyrd-Hansen, 2005). Pearl-growing from these studies identified a key review (Mason *et al.* 2008). This review appears in *Expert Review of Pharmacoeconomics* which does not appear in any of the three databases. To ensure there were no other relevant articles in this journal, a search was performed using the original search terms. No further articles were identified. The review by Mason *et al.* (2008) identifies studies estimating an MVQ that fall into one of the following categories: estimating an MVQ based on VOSL estimates (Abelson, 2003; Hirth *et al.* 2000; Johannesson and Meltzer 1998; Mason *et al.* 2008); estimating WTP through CV with members of the public (Johannesson and Johansson 1997; Johnson *et al.* 1998; Gyrd-Hansen 2003; Byrne *et al.* 2005; Donaldson *et al.* 2008); estimating WTP through CV with patients (Blumenschein and Johannesson 1998; Zethraeus, 1998; Cunningham and Hunt, 2000; King *et al.* 2005). The publication by Donaldson *et al.* (2008) presents preliminary results of the Social Value of a QALY (SVQ) project. This was a large scale project commissioned by NICE in 2004. More recent publications by the research team have now become available (Baker *et al.* 2010). Furthermore five recent publications using CV were identified through a prospective pearl growing approach (Lieu *et al.* 2009; Prades *et al.* 2009; Bobinac *et al.* 2010; Shiroiwa *et al.* 2010; Zhao *et al.* 2011). One further article was identified from the Health Economics Study Group archive (Mason and Donaldson, 2007). The results of the large scale European Value of a QALY (EuroVaQ) have also recently been made available through an online report (Donaldson *et al.* 2011; www.research.ncl.ac.uk/eurovaq), and a paper was also presented at the recent joint HESG/CES conference in Aix en Provence (Robinson *et al.* 2012).

These studies will be reviewed before a discussion of the broader issues in estimating a 'one-size fits all' MVQ are presented. In reviewing existing studies estimating an MVQ the following information is extracted: the country in which it was performed, the sample size, whether the study uses patients or general population, the outcome being valued, the WTP elicitation method, the payment vehicle, whether any discounting was applied, and of course the MVQ estimates. Furthermore, in studies directly eliciting preferences for health states the method used will be noted, as will the method of combining the quality of life data and the WTP values.

Table 2.2 outlines the 20 studies estimating an MVQ that were identified in the review. These studies will now be discussed in more detail in the following sections.

Table 2.2 - Summary of existing MVQ estimates in the literature (all in 2009 prices)

Study	Country	Sample Size	Approach	Elicitation Method	Payment Vehicle	Discount Rate	Mean MVQ Estimate
Estimates Based on the Value of a Prevented Fatality (VPF)							
Johannesson and Meltzer (1998)[1]	Sweden		VPF (Simple)			3%	**£66,000**
Hirth *et al.* (2000)	US		Review of value of life estimates used to infer MVQ			3%	**£119,000**
Abelson (2003)	Australia		Value of Statistical Life Estimated by observing those in other countries			3%	**£60,000**
Mason *et al.* (2009)	UK	167	VPF (Simple)	Payment Cards	Out of Pocket	1.5%	**£73,000**
		167	VPF (Complex, Including 18-40 yr olds)	Payment Cards	Out of Pocket	1.5%	**£36,000**
		167	VPF (Complex, Excluding 18-40 yr olds)	Payment Cards	Out of Pocket	1.5%	**£70,000**
Studies Eliciting Values from the Public							
Johannesson and Johnson (1997)	Sweden	2824	WTP for increase in life expectancy	Dichotomous Choice	Insurance Premium	0%	**£1,000-£2,000**
Johnson *et al.* (1998)	Canada	246	WTP for increased life expectancy	Conjoint Analysis	Out of Pocket	0%	**>£8,000**[2]
Gyrd Hansen (2003)	Denmark	3201	WTP for change in health status (EQ-5D)	Dichotomous Choice	Out of Pocket	0%	**£7,000 & £8000**
Byrne *et al.* (2005)	US	193	WTP for alleviation of osteo-arthritis	Open-Ended Question	Out of Pocket	3%	**£900-£4,000**
Mason and Donaldson (2007)	UK	52	WTP for change in health status (EQ-5D)	Open-Ended Question	Out of Pocket	0%	**£5,000**
		52	WTP for Life extending treatment	Open-Ended Question	Out of Pocket	0%	**£2,000 & £6,000**
		52	WTP for Life Saving Treatment	Open-Ended Question	Out of Pocket	0%	**£19,000 & £25,000**
Baker *et al.* (2011) SVQ	UK	403	WTP for alleviation of stomach problems and head aches	Random Card Sorting	Out of Pocket	0%	**£18,000 - £700 million**

Table 2.2 (Continued)

Prades *et al.* (2009)	Spain	892	WTP for more effective treatment to alleviate symptoms described by EQ-5D	Random Card Sorting	Out of Pocket	0%	**£4,000-£110,000**
Donaldson *et al.* (2011) EuroVaQ	9 European Countries	21,965 (UK:2,312)	WTP to either avoid risk of EQ-5D health state or avoid given time in EQ-5D health state	Random Card Sorting	Out of Pocket	0%	**All: £7,000-£33,000 UK:£8,000-£18,000**[3]
Shiroiwa *et al.* (2010)	6 countries	5520	WTP for one additional year of life in full health	Dichotomous Choice and Bidding Game	Out of Pocket	0%	**£22,000 (UK) - £47,000 (Taiwan)**
Bobinac *et al.* (2010)	Netherlands	1091	WTP to avoid moving from one EQ-5D state to a worse one	Payment Scale and Open-ended	Out of Pocket	0%	**£8,000-£21,000**
Studies eliciting values from patients							
Blumenschein and Johannesson (1998)	US	69	WTP for alleviation of asthma symptoms	Dichotomous Choice and Bidding Game	Out of Pocket	0%	**£5,000-£35,000**
Zetheraeus (1998)	Sweden	104	WTP to continue hormone replacement therapy	Dichotomous Choice	Out of Pocket	0%	**£9,000-£12,000**
Cunningham and Hunt (2000)	UK	40	WTP for dentofacial surgery	Payment Card	Out of Pocket	0%	**£600**
King *et al.* (2005)	US	391	WTP for improvement from current health to perfect health	Bidding Game	Out of Pocket	3%	**£9,000-£23,000**

Table 2.2 (Continued): Studies eliciting values from patients and the public

Lieu *et al.* (2009)	US	952	WTP to avoid herpes zoster scenarios valued through TTO	Bidding Game and Open-Ended	Out of Pocket	3%	**Public: £20,000 Patients: £23,000-£43,000[4]**
Zhao *et al.* (2011)	China	632	WTP for improvement from current health to perfect health as specified by EQ-5D and SF-6D	Bidding Game	Out of Pocket	0%	**Patients: £4,000 Public: £3,000**

[1]The source of the VPF data was not clear in these two studies
[2]No quality adjustment available for 'minimal activity restrictions'
[3] Top 1% Trimmed
[4] These are trimmed means - highest and lowest 2.5% of values excluded

2.3.2 Studies modelling an MVQ from the existing 'value of preventing a statistical fatality' currently used in public sector safety policy

The simplest way of estimating an MVQ based on the existing roads/rail VPF is to divide the VPF by the discounted QALYs gained for a saved statistical life. This is the approach adopted by Johannesson and Meltzer (1998), Hirth *et al.* (2000) and Abelson (2003). Johannesson and Meltzer (1997) combine a VPF of £1.3 million pounds with an average number of life years lost through a traffic death in Sweden of 30.5 years and rough QALY weights in the Swedish general population (Brooks *et al.* 1991). Abelson (2003) use a VPF of £1.7 million pounds and assume 40 years of remaining life expectancy. This value is less precise than the others as Australia has no general VPF and so one was derived using overseas values rather than through a direct survey of the public. This study also assumes that one life year is equivalent to one QALY, which is likely to underestimate the value of a QALY by overestimating population health (at perfect health). Abelson (2003) produces an MVQ estimate of £60,000 compared with £66,000 in the study by Johannesson and Meltzer (1998). Both Studies employ a discount rate of 3%.

Hirth *et al.* (2000) identify 37 articles that present estimates of the VPF and use these to infer the value of a QALY. Eight of these studies asked either WTP for a risk reduction or WTA risk increases (the other studies used human capital or revealed preferences). These studies came from France, Canada, the UK, Denmark and the USA. For each of the studies, remaining life expectancy was calculated for the study population using US life tables and quality-of-life adjustments were made using age-specific weights from the Beaver Dam Health Outcomes Study (Fryback *et al.* 1993). A 3% discount rate is used. No mean MVQ is presented in this study. The median MVQ based on the eight CV studies was £119,000 (all other values are means unless stated otherwise). However, it should be noted that there is wide variation in the VPF estimates from the eight different CV studies, ranging from $1,230,828 (1997 US$) to $25,926,349. The authors acknowledge this is a result of differing methods, study populations and data sources. The study with the lowest estimates considers road safety (Desaigues and Rabel, 1995), while the study with the highest estimates considers job risk (Lanoie *et al.* 1995). Furthermore, some of the studies use WTP while others use WTA.

More recently, Baker *et al.* (2008) has questioned the simplistic approach above and focused on the way in which the VPF varies with the age of those who will enjoy a reduction in the risk of premature death. Clearly respondents who have longer to live should be prepared to pay more to reduce the risk of death. Therefore, Baker *et al.* (2008) argue estimates of the value of a QALY should be based on the rate at which the VPF increases with increasing life expectancy.

Mason *et al.* (2009), estimate MVQs through the simplistic approach (assumes each future life year is of equal value), and a more 'sophisticated' approach that accounts for the inverted U-Shaped relationship between the VPF and age, using the current UK VPF derived by Carthy *et al.* (1999). In this sophisticated approach the rate at which the VPF increases with life expectancy will be negative in the early years of life (hence yielding a negative value of a life year). The point at which the rate shifts from being negative to being positive is found to be at approximately age 40. Therefore the authors adopt two approaches. The first is to include all respondents and simply assign negative values to those aged between 18 and 40. The second is to exclude respondents between the ages of 18 and 40 from the analysis. This thesis takes the view that neither approach is preferable to the simplistic approach. Negative values are fundamentally flawed as they suggest that road traffic accidents are a positive occurrence. The exclusion of such a large demographic is equally as indefensible and is likely to bias results. Provided data collection utilises a sufficiently representative sample of the population differing WTP values between the age groups should already capture the inverted U-shaped age relationship. Furthermore, given that these approaches still result in one MVQ to be applied universally it will not result in differing allocation of resources aimed at different demographic groups, which seems to have been Baker's initial motivation (Baker *et al.* 2008). The MVQ estimate for the simplistic approach (with discounting) is £73,000. The MVQ estimate for the 'sophisticated' approach, including 18-40 year olds, is £36,000. The MVQ estimate for the 'sophisticated' approach, excluding 18-40 year olds, is £70,000. This shows that the effect of excluding the 18-40 year olds is large.

A further point of controversy arising from the study by Mason *et al.* (2009) is the adoption of a 1.5% discount rate, rather than the 3% rate favoured by other VPF based MVQ studies. In fact the discount rate recommended in Her Majesty's Treasury Green Book is 3.5% (Great Britain HM Treasury, 2004). Their justification for using a 1.5% discount rate is that the full discount rate of 3.5% is made up of two parts: a rate of pure time preference (1.5%) and a part that reflects diminishing marginal utility of consumption (2%). The authors argue that the WTP-based VPF can be expected to grow at much the same rate as the marginal utility of consumption declines and so to include this part would be a form of double counting. This suggests that the previous estimates, applying a discount rate of 3% have under-estimated the true MVQ.

Across all VPF based studies the MVQ estimates range from £36,000 to £119,000 (Note, the £119,000 figure is a median while others are means, so caution is warranted in comparing the results). However, excluding the low value generated by the rather questionable approach taken by Mason *et al.* (2009) to exclude respondents aged 18-40, estimates range from £60,000 to £119,000. The broad range of values generated led Hirth (2000) to conclude "because of the considerable inter-study variability and methodologic questions regarding issues such as whether individuals tend to overestimate small risks, the value of life literature can provide only a rough empirical basis for a decision rule for CEA" (p.340). Mason *et al.* (2009) also acknowledge the lack of evidence looking at whether individuals' preferences for road safety are the same as their preferences for health care. As will be seen in the next section, MVQs based on a health context are much lower than those that have been inferred from the VPF.

2.3.2.1 Summary of the VPF based approach to estimating an MVQ

There are a number of difficulties with this approach. Firstly it relies on the assumption that VPF estimates are accurate and consistent, and a number of reasons why this might not be the case have already been outlined. Indeed, there are a wide range of available VPF estimates with no clear guidance to researchers which is preferable. Secondly, the VPF is based on the reduction of the risk of a road accident. It is unclear whether individuals' preferences for safety are the same as their preferences for health care. Chilton *et al.* (2002)

estimate preference based values of safety in three contexts – rail, domestic fires and fires in public places and compared them with the values for road safety. The results imply the VPF is fairly constant regardless of the cause of death respondents are faced with. However, Tolley *et al.* (1994) show that WTP to reduce the risk of death varies with the type of death. WTP to avoid an unforeseen instant death was $2 million, compared with $2.75 million for avoiding death by heart disease and $4 million for death by lung cancer (1994 US$). Average WTP to reduce risk of death in the London underground is 50% higher than WTP to reduce road fatalities (Jones Lee *et al.* 1998). While these studies do not directly compare preferences for health and road safety it seems likely that the two will be different.

The Dft report mentioned earlier (Department for Transport 2011) takes a very damning view of the prospect of inferring the value of a QALY from the VOSL. The strongest objection is that the VOSL is a risk based concept, while the QALY is not. This is particularly so in the NICE context, where the TTO method of quality adjustment is preferred over the SG approach. The report concludes that 'in the absence of more empirical evidence it thus seems hard to see why any close relationship should be supposed between the VPF and any monetary valuation of the QALY" (p.36).

2.3.3 Studies directly eliciting a WTP value for changes in health status/increases in life expectancy

Since multiple factors vary across the existing studies identifying the specific causes of differences in MVQ results is difficult. This section will consider different aspects of study design in an attempt to determine their effect on results. More detailed descriptions of the individual studies directly eliciting values can be found in Appendix A**2**. The UK SVQ (Baker *et al.* 2010) and EuroVaQ (Donaldson *et al.*2011) projects are considered in more detail in the subsequent section.

2.3.3.1 What is valued?

Increase in Life Expectancy

Two studies ask members of the general public their WTP for an increase in their life expectancy (Johannesson and Johansson, 1997a; Johnson *et al.* 1998). In the first of these studies (Johannesson and Johansson, 1997a) respondents were told the chance they would survive to the age of 75 (varied depending on the age and sex of the respondent) and then told that upon reaching 75 they could have a treatment that would increase their life expectancy from 10 years to 11 years (i.e. live to 86 rather than 85). Respondents were asked if they would pay a given insurance premium immediately (respondents were randomised to 1 of 6 different values). The approach taken by Johnson *et al.* (1998) is in fact a form of conjoint analysis. Rather than state a WTP value, respondents must chose from two options, which each have four attributes: symptoms, longevity, cost over the next three years and daily activities. A series of pair wise choices were presented to respondents and they were asked to state whether option A or option B was preferred on a 7-point scale, ranging from 'A is much better' to 'B is much better'.

Avoidance of Immediate Death

Shiroiwa *et al.* (2010) ask people to value a treatment that will prevent immediate death and deliver one additional year of life in full health. The study is particularly interesting because as well as eliciting individual WTP, they elicit WTP for a gain of one QALY in a family member, and WTP for a gain of one QALY in a member of society. The individual and family questions use an out of pocket payment method, while the societal question uses an increase in taxation. The results show that in six of the seven countries WTP (family) exceeds WTP for themselves (individual) which suggests high levels of altruistic utility. In three of the countries (Japan, UK and Australia) WTP (society) exceeds WTP (individual) which could be caused either by a sense of altruism or through a form a free riding as they only bear a proportion of the cost through a tax based payment method.

EQ-5D Health Improvements/Reductions

The EQ-5D (Brooks, 1996) has five health dimensions: mobility, self-care, usual activities, pain and anxiety/depression. There are three levels for each dimension: no problems, moderate problems, extreme problems. Therefore there are a total of 243 (3^5) different health states. For example, state 22322 describes "some problems in walking about, some problems with self-care, unable to perform usual activities such as work and study, moderate pain or discomfort, moderate anxiety or depression".

Four studies ask respondents to value health improvements as specified by the EQ-5D (Gyrd-Hansen, 2003; Byrne *et al.* 2005; Prades *et al.* 2009; Zhao *et al.* 2011). The study by Gyrd-Hansen (2003) uses the same 42 health states employed in the derivation of the UK EQ-5D population value set (Dolan, 1997). Respondents were first given a discrete choice between two health states (the decrement between the two health states ranged from 0.012 to 0.312 on the EQ-5D Index, as indicated by the UK value set). They were then asked if they would pay a specified monthly out of pocket payment (to which they were randomly allocated) to have a medicine that would improve their health from the poor EQ-5D state to the better one. The time period over which this will occur is not specified, presumably it is indefinite until death. Given that both costs and benefits occur simultaneously the duration over which this health improvement last is of limited importance (see section XX on discounting). Byrne *et al.* (2005) presented 193 members of the US public with two osteoarthritis scenarios, descriptions of which were based on the EQ-5D domains. Utility values for the health states were elicited through VAS, TTO and SG. WTP to move from each of the two health states to full health, as well as from own health to full health, was elicited through an open-ended lump sum question. Prades *et al.* (2009) use a complex study design (see Appendix A4 for more detail). In basic terms they ask respondents their WTP to have a more effective treatment (works more quickly) to improve their health from an impaired EQ-5D health state to perfect health. Zhao *et al.* (2011) ask both chronic prostatitis patients and general population to value their own health through EQ-5D and SF-6D. They then elicit WTP for an improvement from current health to perfect health. They find slightly higher MVQ values for patients than for members of the general public.

Bobinac *et al.* (2010) ask respondents their WTP to avoid moving from one EQ-5D health state to a worse one (as determined through a pairwise choice). The health states were valued through VAS. A total of 42 EQ-5D health states were used and paired into 29 choice scenarios. The majority of the pairs were used in deriving the UK EQ-5D value set.

Disease Specific Health States

Baker *et al.* (2010) ask respondents to value two specific health states: stomach bowel problems and recurrent episodes of head pain. Respondents were asked their WTP to either prevent the certainty of the given illness or to eliminate some risk of the illness (either 10% or 5%). Following the WTP part of the questionnaire, values for the health states were obtained through SG.

Lieu *et al.* (2009) describe a series of scenarios based on herpes zoster to both a general population sample and patients with either shingles or postherpetic neuralgia. The health states are valued through TTO and WTP elicited for an improvement from the given health state to perfect health.

Own Health

Five studies ask patients to value their own health (Zetheraeus, 1998; Blumenschein and Johannesson, 1998; Cunningham and Hunt, 2000; King *et al.* 2005; Zhao *et al.* 2011). Blumenschein and Johannesson (1998), and King *et al.* (2005) ask patients to value their QoL through VAS, TTO and SG and elicited WTP for an improvement from current health to perfect health. Blumenschein and Johannesson (1998) use asthma sufferers, while King *et al.* (2005) use a combination of general medical patients, patients with a degenerative spine condition, and patients with cerebral aneurysms. Zetheraeus (1998) ask women their WTP to continue receiving Hormone Replacement Therapy (HRT). Current health with HRT is assessed through EQ-5D, while QoL without HRT is assessed through VAS and TTO. Cunningham and Hunt (2000) ask people with dentofacial deformities their WTP to receive treatment to correct these deformities. Their state of health is assessed through SG.

Do results differ depending on what is valued?

The two studies valuing increases in life expectancy produce MVQ estimates ranging from £1,000 to >£8,000, (Note: no quality adjustment is available for 'minimal activity restrictions' in the study by Johnson *et al.* 1998 which means the estimate of £8,000 is likely to be an underestimate as it is the value of a year in less than full health). The study valuing a life saving treatment produces an estimate of £22,000 in the UK (Mason and Donaldson, 2007). The four studies that specify health improvements using the EQ-5D produce MVQ estimates ranging from £900 to £110,000. The result produced by Prades *et al.* (2009) of £110,000 is an outlier, the other three studies produce results in the range of £900 to £21,000. There is no obvious explanation for why the Prades *et al.* (2009) study produces such high estimates. The five studies asking patients to value own health produce estimates ranging from £3,000 to £35,000. It seems that the studies valuing increases in life expectancy and the studies using the EQ-5D produce similar estimates (with the exception of Prades *et al.* 2009).

Mason and Donaldson (2007) directly test whether results differ depending on whether respondents value QoL improvements, life extensions or lifesaving treatments. To clarify, a life extending treatment is one that adds a specified amount of time onto the end of an individual's life expectancy, while a lifesaving treatment is one that prevents imminent death. They find mean MVQs of £5,000 for the QoL improving question and £2,000 for the life extending question. These values are similar to those presented above, but suggest that QoL improving treatments may produce higher estimates (as found by Prades *et al.* 2009). In the two lifesaving questions the MVQ estimates were £19,000 and £25,000. As seen in Table 2.2 the VPF based studies produce higher estimates than life-extending or QoL improving studies.

2.3.3.2 Elicitation Method Used

Of the two studies valuing increases in life expectancy one used a discrete choice elicitation method (Johannesson and Johansson (1997) and the other used a form of conjoint analysis (Johnson *et al.* 1998). The conjoint analysis approach produced the higher results.

Of the four studies valuing health improvements described by the EQ-5D one uses a discrete choice approach (Gyrd-Hansen 2003), one uses an open ended question (Byrne *et al.* 2005) and one uses a random card sorting approach (Prades *et al.* 2009). Bobinac *et al.* (2010) use a payment card approach followed by a bounded open-ended question (the bounds being determined by the results of the payment card exercise). The authors argue this approach adds robustness to their results. The highest results occur with the use of the random card sorting approach, and the lowest estimates occur with the open-ended question. Gyrd-Hansen (2003) acknowledges that the elicitation method could have influenced the results. Of respondents presented with the highest amount (10,000DKK=£1,350) 15.9% were willing to pay more than this amount for the health improvement, which suggests that the choice of values was not appropriate to capture the maximum WTP. Furthermore, 17.9% of respondents presented with the smallest amount (100DKK=£13.50) refused to pay it. Byrne *et al.* (2005) argue that using an open-ended question may have biased results downwards (although no explanation is given for this argument).

Of the five studies eliciting valuations of own health from patients one uses discrete choice (Zetheraeus, 1998), two use a bidding game (King *et al.* 2005; Zhao et al. 2011), one uses discrete choice and bidding games (Blumenschein and Johannesson, 1997) and one uses a payment card format (Cunningham and Hunt, 2000). The study using both discrete choice and bidding game formats found that the bidding game format produces considerably lower results. However, Zetheraeus (1998) uses discrete choice to produce estimates ranging from £9,000 to £12,000, while King *et al.* (2005) using a bidding game format and produce estimates ranging from £9,000 to £23,000.

In contrast to the above studies showing a wide variability in results depending on the elicitation method used (although this effect cannot be isolated from other confounding factors), Shiroiwa *et al.* (2010) use both a double bound dichotomous choice method (analysed by a non-parametric Turnbull method) and a bidding game format. They find very similar results in all six countries studied.

2.3.3.3 Payment Vehicle

All studies except one use out of pocket payment. Johannesson and Johansson (1997) use an insurance premium as the method of payment. Interestingly this study produces the lowest estimates of all the studies.

2.3.3.4 Rate of Time Preference

Only three of the studies apply discounting (Byrne *et al.* 2005; King *et al.* 2005; Lieu *et al.* 2009). All three of these apply a discount rate of 3%, which is the rate recommended by the US Panel on Health in Cost Effectiveness in Health and Medicine (Gold *et al.* 1996). In some cases not applying discounting should not be problematic. For example, in the study by Gyrd-Hansen (2003) payment and health improvements occur simultaneously (through on-going monthly payments) and public economic evaluation guidelines now often recommend that costs and health benefits be discounted equally (e.g. NICE, 2008). Similarly, no discounting is applied in the studies by Zetheraeus (1998) and Blumenschein and Johannesson (1998) since the costs and benefits are incurred at the same time. However, in the two studies eliciting WTP for life extensions (Johannesson and Johnson 1997; Johnson *et al.* 1998) no allowance is made for time preference even though the authors acknowledge the importance of time preference. Given that payment is made in the present for a benefit that occurs a number of years in the future, if a positive discount rate were applied to these benefits the MVQ estimates would become lower.

While Byrne *et al.* (2005) apply a discount rate of 3% to the baseline results, sensitivity analyses apply discount rates of 5% and 1%. The results show that the effect of varying the discount rate is close to £1,000. Since the baseline results range from £900 to £4,000 this is a significant variation caused by varying the discount rate.

Shiroiwa *et al.* (2010) attempt to estimate a rate of time preference by not only eliciting respondent WTP for a lifesaving treatment that delivers one QALY immediately, but also WTP for a lifesaving treatment that delivers one QALY in five years time. They estimate implicit discount rates of 6.8% in Japan, 3.7% in South Korea, 1,6% in Taiwan, 2.8% in the UK, 1.9% in Australia, and 3.2% in the US. These results suggest that the use of a 3% discount rate in the three studies above was appropriate as they were all conducted in the US.

2.3.3.5 Method of eliciting utility values

Studies eliciting utility values through more than one preference elicitation method typically find that the VAS produces the lowest values, SG produces the highest estimates and TTO produces intermediate values (King *et al.* 2005; Blumenschein and Johannesson, 1998). This clearly has implications for the calculation of an MVQ. For example, a higher utility value generated through the SG method would produce lower MVQ estimates than an MVQ estimated using a lower utility value generated through VAS. This suggests that comparison cannot be made across methods. The SG method has been referred to as the gold standard of preference elicitation (Torrance and Feeney, 1989) since it is strongly rooted in von Neumann-Morgenstern expected utility theory (von Neumann and Morgenstern, 1944). However, this position has been challenged (Richardson, 1994) and recently some have sought to place TTO in a theoretical framework (Buckingham and Devlin, 2006). Further examination of the preference elicitation methods, with a focus on TTO, is presented in the next chapter.

Bobinac *et al.* (2010) estimate utilities through either the Dutch EQ-5D tariff (which is derived through TTO valuations) or through direct VAS calculations. The difference in the mean utility values is only small, 0.32 based on the tariff and 0.33 based on VAS valuations, but is statistically significant owing to large variations at the individual level. Owing to the calculation method (see next section) this drives large differences in the MVQ estimates. Using the tariff based utilities gives an MVQ of £21,000, while using the VAS valuations gave an estimate of £8,000.

2.3.3.6 Method of combining WTP values with utility values

Baker *et al.* (2010; SVQ) raise an interesting point regarding the method of calculating the MVQ, which is relevant to the calculation method used in the following chapter. When a study elicits both a WTP for a health improvement *and* a utility value for the given health improvement, how should the information be combined to generate an MVQ estimate? The approach more closely aligned with welfare economic theory, since it allows for individual variation in results, would be to estimate an MVQ for each individual and then calculate an average of the individual estimates (the 'individual approach'). However, in the SVQ study since some respondents are only prepared to accept a very small amount of risk in the SG exercise they generate extremely high MVQ estimates. A total of 115 respondents (out of 403) give answers that generate estimates of more than £1 million, and some of the estimates are thousands of millions of pounds. This leads to mean MVQ estimates of £300 million for the sample valuing stomach problems, and £700 million for those valuing headaches. Therefore, Baker *et al.* (2010) propose an alternative which is to calculate the mean WTP and the mean QALY gain and combine the two (the 'aggregate approach'). This approach is similar to the approach used in the VPF studies, which make no attempt to estimate individual based MVQs. An alternative would be to exclude extreme outliers following clearly set criteria, but there are no strong grounds on which to base this.

If one wishes to subscribe to the conventional precepts of welfare economics and if one could be confident that individuals have values and preferences that conform with standard assumptions (such as diminishing marginal utility, Arrow 1951), and that their responses reveal those values with total accuracy and precision, the 'individual approach' would be appropriate. However, if respondents are not able to state their preferences with total accuracy, these inaccuracies will be magnified under the 'individual approach'. This is because two statements of preference from an individual are multiplied together. Inaccuracies in either of the stated preferences will therefore be magnified. The 'aggregate' approach can alleviate this problem because each stated preference is aggregated across individuals before it is combined with the other preference. This aggregation should work to minimise the 'noise' surrounding individual preferences prior to multiplication. The distinction between the 'individual' and 'aggregate' approaches becomes more important when very small trades are permitted in the preference elicitation exercise. The problem could be limited by increasing the smallest unit of trade allowed in the exercise but this can result in an increased number of 'non-trades'.

Six of the existing MVQ studies have combined utility values with WTP results using the 'individual' approach (Byrne *et al.* 2005; King *et al.* 2005; Lieu *et al.* 2009; Bobinac *et al.* 2010; Zhao *et al.* 2011). Four studies have used the 'aggregate approach' (Blumenschein and Johannesson, 1998; Zetheraeus, 1998; Cunningham and Hunt, 2000; Prades *et al.* 2009).

The studies using the 'individual approach' do not produce particularly high results (they range from £900 to £47,000) suggesting very high individual values is often not a problem. In the Byrne *et al.* (2005) study the highest mean utility value was 0.858 which perhaps suggests that the effect of very small utility decrements was of limited importance. However, Lieu *et al.* (2010) encounter considerable difficulties. Of 8099 responses available for analysis, in 3033 of them (37%) respondents had traded no time in the TTO exercise. The number of 'non-traders' is fairly constant across patients and general population: 33% of the general population sample, 41% of the shingles patients and 40% of the postherpetic neuralgia patients 'non-traded'. Since these responses result in an undefined MVQ in the individual approach, the authors exclude these responses. Even when these responses have been excluded very high values for some individuals skewed the results. These high MVQs were primarily driven by very low amounts of time being traded in the TTO exercise. The authors address this by presenting trimmed means, in which the highest and lowest 2.5% of results were excluded. They also present medians. Note, the TTO exercise is used in Lieu *et al.* (2010) values a temporary mild health state rather than a chronic state followed by death, as in the conventional TTO. For example, one of the scenarios asked respondents how much time at the end of their life they would be willing to give up to avoid getting the flu for one week now. This is obviously likely to lead to a higher prevalence of non-trading than conventional TTO valuations of EQ-5D states, for example.

In the study by Zhao *et al.* (2011) some respondents value their current health as perfect health. Of 268 chronic prostatitis patients, 16 value their health as perfect health on the EQ-5D, while 169 of 364 in the general population sample valued their health as full health. These respondents were excluded from the analysis.

2.3.3.7 Country in which the study is performed

The existing studies allow us to compare results between different countries both between different studies, and within a study. Three CV based studies have been performed in the UK (Cunningham and Hunt, 2000; Mason and Donaldson, 2007; Donaldson *et al.* 2008). They produce results ranging from £6,000- £25,000 (excluding the very high outlier in the Donaldson *et al.* 2008 study). Four studies have been conducted in the US producing results ranging from £900 to £43,000. Shiroiwa *et al.* (2011) elicit MVQ estimates in six countries and find that the results increase in the following order: UK, Japan, Australia, US, South Korea, Taiwan. The authors find a strong relationship between the proportion of expenditure on health that is privately funded and the size of the MVQ estimate. Countries in which health care is predominantly privately funded (such as the US) produce higher estimates than countries in which the majority of care is provided by the state.

2.3.4 The UK Social Value of a QALY (SVQ) and the European Value of a QALY (EuroVaQ) projects

In light of various pressures on NICE to either raise (Lakhani, 2008) or lower (West, 2009) the cost effectiveness threshold, in 2004 NICE co-funded a research project to consider the value of a QALY, the UK Social Value of a QALY (SVQ) project. Preliminary results from this project are now available (Baker *et al.* 2010). The European Commission has also instigated a project to consider the value of a QALY across 9 European countries (UK, Spain, Denmark, Sweden, Poland, Hungary, Netherlands, Norway, France), the European Value of a QALY (EuroVaQ) project. The results of this project have also recently become available (Donaldson *et al.* 2011).

The SVQ project conducted interviews with a convenience sample of 409 members of the general public. There were four versions of the questionnaire. Versions 1 and 2 focused on scenarios about stomach illness, while versions 3 and 4 focused on scenarios about head pain. Within each version there were three different durations of illness: 3 months, followed by a return to respondent's current state of health; 12 months followed by return to respondent's current state of health; and a chronic condition where the illness lasted for the rest of the respondent's life. The key way in which Version 1 differed from Version 2, and likewise the way in which Version 3 differed from Version 4, related to the questions involving the risk of an illness: Version 1 and 3 asked about WTP to eliminate 10% risks of the three durations of illness, while Version 2 and 4 asked about eliminating 5% risks. Respondents were asked five WTP questions. For example, in version 1 respondents were asked: their WTP for a 'simple, safe and painless cure that would avoid' the certainty of the 3 month stomach illness; their WTP to avoid the certainty of the 12 month stomach illness;

their WTP to eliminate a 10% risk of the 3 month stomach illness; their WTP to eliminate a 10% risk of the 12 month stomach illness; their WTP to eliminate a 10% risk of suffering the stomach illness for the rest of their lives. A random card sorting procedure was used and payment was out of pocket. The results of the WTP exercises showed that the values given for the 12 month duration of illness was considerably less than four times the value for the 3 month duration (2.3:1 for stomach illness and 2.2:1 for head pain). The most likely explanation offered for this is the influence of the budget constraint. As the duration of the illness increases the budget constraint becomes more influential. Another worrying finding is that there were no statistically significant differences in the WTP values for 10% risk and 5% risk of the illnesses.

Following the WTP questions in the SVQ project respondents were asked to value the health states through Standard Gamble (SG). SG is preferred to TTO as it comes from the same theoretical stable as WTP, but Donaldson *et al.* (2011) acknowledge that since the QALY tariff used by NICE is based on TTO it is not necessarily the case that direct comparison between the value of a QALY derived from SVQ and NICE's valuation of a QALY can be made. The choice for SG over TTO is surprising given the project was co-funded by NICE, with the aim of addressing the appropriateness of the current cost effectiveness threshold. There were a number of different SG questions asked: 3 months for certain vs a gamble with outcomes of return to current health or 12 months in the state; 12 months for certain vs a gamble with outcome of return to current health or rest of life in the state; and rest of life for certain vs gamble with outcomes of current health or immediate death). One question was included to explore the relationship between the subjective loss entailed by the 3-month illness compared with the loss of well-being associated with suffering the condition for four times as long (12 months). Respondents were asked to choose between the certain prospect of 3 months in either of the two conditions and a treatment whose outcome was uncertain between current health or 12 months in the condition. If the 12 month illness involves four times as much QALY loss as the 3 month illness, a respondent behaving broadly according to the QALY model should feel that the two alternatives are evenly balanced when the risk of failure is 0.25. However, the majority of respondents were willing to accept only a smaller risk. On average, the 12 month illness appeared to be regarded as five or six times as bad as the 3-month duration. When viewed in relation to the responses to the WTP questions which were insensitive to duration, the results suggest either that budget constraint effects are really quite strong or that framing questions in terms of 'chances' may prompt excessively cautious responses.

As well as concerns arising from the raw WTP and SG responses, combining these results to estimate a QALY value was also problematic, as outlined in section 2.3.3.f above. Essentially, estimating a QALY value for each individual and then aggregating can lead to a very high mean value due to the influence of very small trades. Aggregating the WTP and SG responses, prior to combining them, on the other hand, is theoretically questionable.

The EuroVaQ project attempts to address a number of the problems encountered in the SVQ project. One factor possibly contributing to the extremely high values estimated previously is, as the duration and/or risk of the health gain presented in the WTP questions was 'fixed' in advance, respondents were potentially each valuing very different QALY gains. So, for example, respondents who had attached a very high utility value for the health state in the standard gamble component would be valuing a very much smaller fraction of a QALY than those who had attached a lower utility value to the health state. The EuroVaQ study seeks to employ a study design that allows all respondents to value the same sized QALY gain. This was done through the development of two different types of WTP questions: 'risk variant' and 'time variant' questions. Risk variant WTP questions ask respondents about their WTP to avoid some risk of a health state (that they had previously given a utility value for). The WTP questions presented to individual respondents were customised according to their utility valuation by varying either risk or time in order to keep the QALY gain constant across respondents. So for example, setting the QALY gain at 0.05 for all respondents would mean that a respondent whose utility value for a given health state was 0.90, would be asked their WTP to avoid a 50% chance of that health state for one year. Someone who gave a utility value of 0.80 for the same health state would be asked their WTP to avoid a 25% risk of the health state for a year. In time variant WTP questions, respondents are asked about their WTP to avoid some duration of a given health state (again that they have previously given a utility value for), this time with the duration varied in order to keep the QALY gain constant across respondents. For example, for the respondent with a utility value of the health state of 0.90, avoiding 6 months in that state with certainty, would amount to a gain of 0.05 QALYs. Note that this approach clearly relies heavily on the assumptions of the QALY model (such as linearity with risk and duration as outlined in the introduction), and unfavourable results may not be a reflection of the WTP questions but instead suggest that the QALY model does not hold

In the EuroVaQ project all respondents valued two EQ-5D health states (21121 and 22222) through either TTO or SG. The WTP exercise was through a random card sorting procedure as in the SVQ project. Respondents completed a utility elicitation exercise for the first EQ-5D state, and then completed a WTP exercise for the same state. They then did likewise for the other health state. Respondents valued QALY gains of 0.05 and 0.1 in the WTP exercise. So for example, in the risk variant WTP question, a respondent who attached a utility value of 0.8 to state 22222 in the preference elicitation exercise would be presented with a 25% chance of that health state in a 0.05 QALY gain question (Utility decrement of 0.2 * chance of occurring 0.25 = 0.05). In the corresponding time variant WTP question, respondents would be asked their WTP to avoid 3 months of state 22222.

The EuroVaQ project achieved a total of 21,965 internet based interviews with members of the general public, of which 2,312 were conducted in the UK. The results show that the novel methods seem to have eliminated the effect of very small trades in the utility elicitation exercises on the mean WTP per QALY figures. Indeed the estimates reported for the UK are all within a range that NICE generally considers to be cost effective. However, there are still a number of problems. Firstly, the WTP results are insensitive to the size of the QALY gain, so that the WTP per QALY estimates based on smaller QALY gains (0.05 QALY) are higher than those based on larger QALY gains (0.10 QALY). Secondly, the estimated WTP per QALY is not independent of the framing used in the WTP question in that the risk variant format appears to yield higher estimates than a time variant format. The authors had expected respondents to pay more to avoid the certainty of some duration of a health state than to eliminate the risk of that health state. They argue the results might suggest that introducing risk into the WTP questions has resulted in risk aversion influencing the amount respondents are WTP. Thirdly, there were a number of non-traders in both the utility elicitation and WTP exercises. Between 21% and 33% of respondents refused to pay any amount of money for the health gain, depending on the particular question. The exact number of respondents who 'non-traded' in the SG and TTO exercises is not presented. A number of respondents who non-traded in either the SG or TTO exercises were then prepared to pay money to avoid the health state. Conversely, some respondents who traded in the SG/TTO exercises were then unwilling to pay any money to avoid the health states. This clearly does not accord with standard economic behaviour and poses questions over the chained approach.

As well as using the chained approach, the EuroVaQ project also estimates WTP per QALY through a direct approach. There were multiple variants of the direct approach. Respondents were asked (depending on the version they were assigned to): their WTP for a 25 point drop in health (on the 100 point VAS scale) for four years which would happen either in 12 months time or later in life, with payment being either now in one lump sum or in four annual instalments; their WTP for X months of additional life at the end of life or to avoid X months of coma now, with X being determined by their self-reported health so that the gain amounted to one QALY; their WTP to avoid either a 10% or 5% risk of a 10 point drop in health for 10 years. The results again displayed insensitivity to scale. The questions offering smaller health gains generated larger mean WTP per QALY values. Extensions of life were valued, resulting in higher values than improvements in quality of life. Mean responses for increases in quality of life were insensitive to the timing of the gain, while responses for increases in longevity were sensitive to the timing, with increases in the near future being valued more highly than increases at the end of life. It was hypothesised that insensitivity to scale was caused by budget constraints. The inclusion of the questions offering payment in four annual payments rather than one payment allowed this hypothesis to be tested. Indeed, payment on an annual basis generated larger mean values per QALY. Estimates of the WTP per QALY from the direct questions were heavily influenced by a few respondents giving very high values, similar to the problem observed in the conventional chained approach used in the SVQ project. However, although the modified chained approach in the EuroVaQ project largely avoids the influence of outliers, this method still produces higher results than the direct method.

2.3.5 Summary of existing MVQ estimates

Across all studies the estimates range from £600 to £119,000 (ignoring the very high outlier generated in the Baker *et al.* 2010 study). Typically the estimates from the studies directly eliciting utilities and WTP values from either patients or from the general public generate lower estimates than the VPF based studies. The estimates elicited from patients and the public generally fall below the upper bound of the implicit NICE threshold of £20,000-£30,000, while none of the VPF based estimates do. The largest study to date, the EuroVaQ project (Donaldson *et al.* 2011) also produced estimates for the UK that lie within the current NICE threshold. The wide variability in results is likely to be a result of the different outcomes being valued and the different WTP methods employed.

2.3.6 Is it possible to derive one unique MVQ?

If one MVQ exists (and can be derived) CEA can be interpreted in the same way as CBA, since the health benefits can easily be monetised and a clear decision rule applied: if the benefits are greater than the costs the intervention should be introduced. Phelps and Mushlin (1991) argue that if an MVQ exists CBA and CEA will lead to the same decisions (although, this would require flexible budgets). The difference is that CBA explicitly assigns a monetary value to years of life, while CEA does so implicitly. The only other differences they acknowledge between the two approaches are: reporting style, aggregation and multiple costs and benefits. Reporting differs because in CBA the value of life is explicitly stated upfront and then the net benefit is presented. However, in CEA the cost per QALY is calculated and the decision is left unstated. Decision makers must then decide if they are willing to pay this amount. Aggregation differs because CEA is often carried out at a very disaggregated level (maybe an individual patient) while CBA is typically at a societal level. The treatment of non-health benefits also differs between the two methods. In CBA these benefits can be monetised and included in the benefit side of the ratio. In CEA, as outlined in previous chapters, there is some uncertainty over what is included in the QALY. Often non-health benefits will have to be included as negative costs in the numerator of the ratio.

However, the existing studies estimating an MVQ have produced differing results, with estimates varying depending on factors such as the level of risk involved, the size and the type of health gain (e.g. quality of life improving or life extensions). Authors have expressed theoretical concerns about the compatibility of WTP and QALYs. Johnson (2005) argues that "QALYs and WTP make odd bedfellows" (p. 608). The QALY model imposes a number of restrictions on neo-classical theory: it rules out substitution except between the narrowly defined domains of ill-health severity and time, assumes no variation in substitutability across individuals, and requires linearity in time. However, WTP estimates do not impose these restrictions and hence are more consistent with the underlying theory. Johnson argues for a more unified approach using more complex multi-variate stated preference methods to specify a more complete utility function that includes other factors such as treatment context, personal characteristics and income. This line of reasoning leads to a move away from CEA in favour of CBA. Johnson argues this is justifiable because it is the responsibility of economists to identify the most efficient allocation of resources, and the responsibility of philosophers and politicians to preside over issues of fairness.

Attempting to derive an MVQ has been termed as 'building a bridge between CBA and CEA'(Dolan and Edlin, 2002). They argue that some rather restrictive and unrealistic assumptions have to be made to build this bridge. The approach of Johannesson and Meltzer (1998) requires that incomes be held constant across individuals for WTP to be proportional to the QALY gain. Dolan and Edlin relax this assumption and show that health must be additively separable to consumption in the utility function, since the relationship between health and income would influence the ability of an individual to enjoy consumption. Ultimately, Dolan and Edlin argue that it is not possible to link CBA and CEA if: (i) the axioms of EU theory hold; (ii) the QALY model is valid in a welfare economic sense; and (iii) illness hinders the ability to enjoy consumption. This 'impossibility theorem' would suggest a single MVQ does not exist. It also suggests that CEA is not justifiable on strictly welfarist grounds, but only on non-welfarist grounds where the output of health care is judged according to its contribution to health itself, rather than according to the extent to which it contributes to overall welfare (as determined by individual preferences).

Gyrd-Hansen (2003) argues that since CEA and CBA are based on two very different normative perceptions of the role of health versus other goods in society, the task of performing a linear transformation from QALYs to WTP is theoretically unattainable. CBA is based on welfarism which measures health through individual preferences for health outcomes relative to other goods. CEA, on the other hand, is based on the extra-welfarist notion of maximising health and not welfare. CEA also ignores variation across income/social groups. Applying one MVQ would entail overriding individual preferences such as diminishing marginal utility of health. Brouwer *et al.* (2008) compare the search for a social value of the QALY to the search for the holy grail. They argue that one value would ignore any equity considerations.

Smith and Richardson (2005) highlight four core issues that need to be resolved before a meaningful societal MVQ can be derived. Firstly, is Societal WTP simply the sum of individual WTP? The existing studies, reviewed earlier, have typically asked individuals their WTP for an improvement in their own health with payment out of pocket (individual WTP). In a tax funded National Health Service as in the UK individual out of pocket WTP may not be appropriate. The question becomes one of how much an individual is prepared to pay for an improvement in someone else's health, with the important caveat that the individual can also benefit from the services that others receive. Clearly, individual WTP can diverge from societal WTP. However, if social WTP is to be used there are uncertainties around whether social WTP can be compared with individual utility loss resulting from taxation. The second issue is whether individual WTP will map directly into social WTP. Social and individual preferences may be driven by quite different motivations. For example, individual WTP is likely to be determined by a simple of assessment of whether the respondent perceives the benefits to outweigh the costs. Social WTP on the other hand may be motivated by a sense of community and obligation towards others. The third issue is whether personal income is the appropriate budget constraint. If the questions are to reflect a tax funded NHS respondents should be told what contributions others in society will be making. All respondents could be presented with a fixed amount or a fixed percentage depending on social policy. Finally, should we adjust for ability to pay? One approach to account for the fact that the preferences of those with greater wealth receive disproportionate weight has been to use equivalence weighting (Bateman *et al.* 2002). One such approach would be to ask individuals their WTP as a percentage of their income, and then apply the average income of society to this percentage. However, Smith and Richardson argue that any attempt to adjust for individual income will result in a value that no longer reflects total societal WTP.

2.4 Discussion and Conclusion

This chapter has reviewed the ways in which the VPF has been calculated, the problems with these methods and the vast array of varying values that exist in the literature. Ways in which the VPF can be used to derive an MVQ have been considered, and studies directly eliciting a WTP per QALY figure through CV have also been reviewed.

The VPF has most commonly been derived through wage-risk studies, but the identification of accurate wage and risk data is problematic, as is the estimation of appropriate statistical relationships to identify an accurate relationship between the two. CV has been used to estimate a VPF, but has been shown to suffer from insensitivity to the size of the risk reduction. To attempt to overcome this problem the current UK VPF was estimated through a chained SG/CV method. However, this approach is not without problems, most notably in combining the responses to the SG and CV exercises.

The VPF has been used to derive an MVQ, either simplistically by dividing by mean remaining life expectancy, or though more complex methods accounting for the inverted U shaped age relationship of the VPF. Deriving an MVQ in this way has been questioned because VPF's are based on valuations of risk reductions in contexts other than health, and valuations can differ by context.

CV has been used to directly elicit a WTP per QALY figure. A chaining approach is now common, whereby respondents first value a given health improvement and then state their WTP for that improvement. Problems have been encountered in combining the responses because very small trades in the health state valuation exercise can lead to very high WTP per QALY figures when combined with the WTP values. This effect can be reduced if worse health states are valued, but the budget constraint can then artificially constrain responses in the WTP exercises as the outcome being valued becomes larger. Non-traders in either the health state valuation exercises or the WTP exercises are also problematic. The more sophisticated approach taken in the EuroVaQ project, which ensures all respondents value the same sized QALY gain has gone some way to limiting the influence of marginal responses, but still encountered difficulties with non-traders which had to be excluded. The EuroVaQ project also found responses varied depending on whether the time variant or risk variant versions of the questions were used. A further concern is that a number of the studies reviewed have found the CV method to be insensitive to scale. Finally, MVQ estimates seem to vary depending on whether lifesaving, life extending or quality of life improving treatments are valued, posing questions over the concept of a one size fits all QALY. This suggests the studies are more likely to produce a threshold range.

Given the difficulties encountered in existing research, this thesis explores a new method based on the TTO method of preference elicitation. Most of the existing studies conducting health state valuations to be linked with WTP valuations, have used the SG method. However, many decision making bodies, including NICE, have a preference for TTO based QALYs. The new method explored in this thesis is potentially more closely aligned with the QALY model, and TTO based QALYs in particular. In order to consider the potential merits of the TTO method for deriving an MVQ, the next chapter considers the strengths and weaknesses of the method relative to CV. This will also enable identification of criteria against which the new method can be tested.

Chapter 3

Feasibility, Reliability and Validity of the Contingent Valuation and Time Trade-Off Methods

3.1 Introduction

The aim of this chapter is to compare the performance of the CV and TTO methods, and to identify tests that can be applied to assess the performance of an exploratory method outlined in later chapters. As outlined in the previous chapter all existing studies estimating an MVQ have done so using CV to estimate the WTP. This has produced widely varying results. The aim of this chapter is to appraise the CV method in terms of its feasibility, reliability and validity to see whether the existing MVQ estimates are likely to be meaningful. The TTO method is also appraised to see if it may represent a viable alternative for producing MVQ estimates.

The feasibility of a technique refers to its acceptability to respondents and can be assessed by factors such as response rates, completion rates, protest responses and ease of completion (Ryan *et al.* 2001). The response rate refers to the proportion of those invited to participate (e.g. sent a questionnaire) that choose to participate. The completion rate refers to the proportion of those who participate that manage to complete the exercise. Protest responses can occur for various reasons. In the context of CV protest responses can occur if respondents do not feel that they should be asked to pay for the good being valued (e.g. members of the general public in the UK may not feel they should be asked to pay for healthcare because they feel it should be provided by the government). Protest zeros need to be distinguished from zero responses that represent a true statement of preference. Ease of completion if often assessed through qualitative feedback, or through the length of time it takes to complete the exercise. Qualitative feedback can be open-ended or on, for example, a five point scale ranging from easy to difficult.

Reliability refers to the precision of the measurement and therefore to unsystematic deviation in the measurement. Reliability can be assessed through intra-rater reliability, inter-rater reliability or test-retest reliability. Intra-rater reliability refers to the correlation between two measurements of the same outcome, taken as part of the same interview (Torrance *et al.* 1986). Inter-rater reliability refers to the consistency in valuations between individuals completing the exercise (Krabbe *et al.* 1997). The most commonly applied method of assessing reliability is through test-retest which examines the correlation between an initial valuation and a second valuation of the same outcome conducted at a later date (Green *et al.* 2000). This correlation is typically measured through the Pearson product-moment correlation coefficient (see Bland and Altman, 1996) to measure the correlation between the first and second valuations. The Pearson coefficient is a poor measure of agreement as it is only sensitive to a linear relationship between two variables (Bland and Altman, 1996). A preferable alternative which allows for complex relationships is the Intra Class Correlation Coefficient (ICC; Koch, 1982). Throughout the chapter correlation coefficients are Pearson's unless specified otherwise, as this is typically what is presented in the articles reviewed. There is not a specific value of correlation coefficient above which a study is considered reliable, but the nearer to 1 the value the more reliable the study.

It could be argued from a practical perspective that what matters is change in mean values not individual variability as measured by correlation coefficients. Change in mean values from one time point to another is also presented where possible.

A scaling method is valid if it accurately measures what it is intended to measure (Froberg and Kane, 1989). Validity can be broken down into internal validity, construct validity and criterion validity. Internal validity refers to the theoretical basis for the method. It is assessed by examining whether the results are intuitive given the theoretical basis of the method. For example, when considering the CV method one would expect responses to have a positive income elasticity of demand. Construct validity refers to whether the method is influenced by factors other than those it is intended to measure (discriminant validity) and whether the results are related to those of similar measures in the way we would expect (convergent validity). Criterion validity refers to a comparison with some given gold standard, which might be whether the results are replicated in the real-world.

In this chapter two reviews are conducted: one for the CV method and one for the TTO method. First the search strategy and findings of the CV review are presented. Then the search strategy and findings of the TTO review are presented. Finally, the two methods are compared and conclusions drawn for the relative strengths and weakness of each.

3.2 Contingent Valuation

The CV method and the different aspects of study design have been outlined earlier in the thesis. This section presents the search strategy used to identify studies that consider the feasibility, reliability and validity of the method, and the results that this produced.

3.2.1 Search Strategy

A review of reviews was initially chosen as the search strategy to identify the relevant literature. These reviews were then updated through a pearl growing approach.

Initially, a review of reviews considering contingent valuation in health care was conducted. These could be systematic reviews or theoretical/methodological discussions of the method in the context of health. The review by Diener *et al.* (1998) was used as the starting point, as this is a comprehensive review published in a well known journal (Health Economics) and is cited extensively (over 300 citations in Google Scholar). A further four reviews were cited by Diener *et al.* (1998), (Gafni, 1991; Morrison and Gyldmark, 1992; Johannesson, 1993; O'Brien and Gafni, 1996). One of these articles (Johannesson, 1993) is a response to another (Morisson and Gyldmark, 1992). A further two articles were identified as part of this debate (Fox-Rushby, 1993; Gyldmark and Morrison, 1993). Searching for articles that had cited Diener *et al.* (1998) through a pearl growing approach identified a further seven reviews (Klose, 1999; Olsen and Smith, 2001; Ryan *et al.* 2001; Ryan *et al.* 2003; Smith, 2003; Sach *et al.* 2007; Smith and Sach, 2010). The most recent of these was by Smith and Sach (2010), and inspection of their references identified a further review (Cookson, 2003). Pearl growing from the other identified articles identified only one further review (Smith and Sach, 2009). Therefore a total of 16 review articles were identified.

Only two of the identified reviews address the issues of feasibility, reliability and validity comprehensively (Klose, 1999; Ryan *et al.* 2003). The other reviews categorise existing health care contingent valuation studies by study design (e.g. Olsen and Smith, 2000; Diener *et al.* 1998), or are short discussions of the weakness of CV (Cookson, 2003) or identify research priorities for the future (Smith and Sach, 2010). While these will be drawn upon, they do not address the core issues of interest here. In order to identify more recent work addressing the feasibility, reliability and validity of the CV method a pearl growing approach was used to identify articles that cite Klose (1999). A total of 204 articles had cited Klose (as of August 2011). Articles were included if they had been designed to test a specific facet of feasibility, reliability or validity. CV studies that do not allow any of these to be studied were not included. Studies only considering the effect of demographics on WTP values were not included, as this is not the focus of the chapter. Inspection of abstracts led to the inclusion of 38 articles.

Since the Ryan *et al.* (2003) review was published in Expert Review of Pharmacoeconomics it has only been cited 10 times. None of these articles were relevant for inclusion.

3.2.2 Findings

3.2.2.1 Feasibility

The issue of feasibility is not covered in the review by Klose (1999) but a number of references are identified in the review by Ryan *et al.* (2003). Thompson *et al.* (1984) ask 184 patients with either osteoarthritis or rheumatoid arthritis their WTP for an alleviation of their symptoms using an Open Ended (OE) format. The completion rate for the WTP questions was only 27%. The authors feel this is caused by the advanced age of the subjects, their limited education, and by the small proportion of respondents working for pay. However, applying the completion rates, disaggregated by educational level, to the US population, similarly disaggregated, suggested that 26% of US adults would complete the WTP questions as posed in this study. The authors also acknowledge that the completion rate is similar to that obtained by Fischer (1979), who use probabilistic improvements in health rather than incremental improvements. Although the authors acknowledge that the completion rate of 27% 'severely limits the applicability of any methodology', they do not believe it establishes the infeasibility of the CV method. Reluctant or perplexed respondents were not 'prodded' for an answer and the authors feel there is scope to increase completion rates.

Donaldson (1998) tested the feasibility of the CV method by asking pregnant women their WTP for their preferred option of two different delivery units. An OE format was used and was administered via a postal questionnaire. A total of 113 women returned the questionnaire (75% response rate), and 102 of these were suitable for analysis. Of these questionnaires 66% of women expressed a preference for one unit over another. WTP values were given by 47 women, giving a completion rate of 69%. Reasons for non-completion included women not knowing what value to place on their preferred option, no response at all being provided, and six respondents (13%) gave zero responses, indicative of a protest response. Reasons for protest responses included "we pay enough taxes already", or "the government should pay".

CV formats other then OE tend to be more acceptable to respondents. Donaldson *et al.* (1997) test the feasibility of CV by asking respondents OE WTP questions, but for respondents who are uncertain, they are given a bidding game format in which the interviewer iterates upwards from zero in increments of £5. In total 300 parents were asked to participate, of which 82 agreed and were asked their WTP for the provision of child health services (response rate of 27%). Of 188 WTP valuations (each parent made more than one valuation) only 25 were deemed invalid (no valuation accompanied by comments like 'the Government should pay' or 'we pay enough already'). Therefore there was a completion rate of 87%. Furthermore, a majority of the sample (64 out of 82) claimed that they found the WTP questions easy to answer. However, there is no indication in the paper of the proportion of valuations that required the bidding game format, making it difficult to determine the feasibility of the OE format.

Ryan (1996) ask men and women their WTP for IVF treatment (for either themselves or their partner) using a DC format administered through a postal questionnaire. The questionnaire was mailed to 700 women and 200 men, and a total of 353 questionnaires were returned (response rate of 39%). Of these 353 responses, 339 provided a WTP value (through yes/no response), a completion rate of 96%. Of the 14 who refused to answer, ten gave a reason. Eight said that there was no limit to what they would pay, one that they do not like paying for things that do not work, and one that WTP had nothing to do with the value of a child.

Phillips *et al.* (1997) elicit the WTP for poison control centres using a bidding game format. They chose a bidding game format after experiencing an 'unacceptable' number of non-responses in the pre-testing when the OE format was used. The bidding game elicits a high completion rate (926/933 = 99%). Reasons for the seven non-completions are not given. Some studies have directly compared the different elicitation methods, and these are addressed in more detail in the section 'convergent validity'. Briefly, Donaldson *et al.* (1995) compared the OE and payment card methods and found that the response rate was higher and the rate of completion was significantly higher for the payment card method. Frew *et al.* (2003) find higher completion rates in the discrete choice questions when compared with the open ended format.

Flowers *et al.* (1997) elicit qualitative feedback from respondents who complete an OE CV question to elicit their WTP to avoid Gaucher disease. The responses are typically favourable to the method. A majority of respondents felt the method was very clear (79%), called for only moderately difficult decisions (84%) and that the decisions they were asked to make were reasonable (83%). Furthermore, more than 61% of respondents felt comfortable or very comfortable using this method for making decisions about their own health care. In contrast to this Donaldson *et al.* (1997) find that a majority of respondents find OE format CV questions difficult to answer. A total of 450 were asked their WTP for their preferred option of two methods of screening for cystic fibrosis. Follow-up showed that 75% of respondents found the WTP questions difficult to answer. Of those who found the questions difficult 50% said this was because they could not put a value on such a thing or because they had difficulty doing so.

Overall, the findings on feasibility are mixed. Using measures such as completion rates to assess the method are difficult because determining what level is acceptable requires a subjective judgement, the review was not able to identify clear guidance on what level is acceptable. Response rates range from 27% to 75%, while completion rates range from 69% to 99%. The qualitative feedback received in the studies by Donaldson *et al.* (1997) and Flowers *et al.* (1997) are contradictory. Given that there is not one consistent methodology for CV and the good being valued differs between studies this variability in findings is unsurprising.

3.2.2.2 Reliability

As mentioned in the introduction there are multiple tests of reliability, but in this section we focus on test-retest reliability as this is the most commonly applied test. No studies assessing the reliability of CV through means other than test-retest were identified. Klose (1999) identifies four studies that study the test-retest reliability of the CV method (Loehmann *et al.* 1979; Thompson *et al.* 1984; O'Brien and Viramontes, 1994; Flowers *et al.* 1997). In the study by Loehmann *et al* (1979) test-retest reliability was considered acceptable as the correlations ranged from 0.82 to 0.95, (Note: Ranges of correlation coefficients are often presented when multiple methodologies are used within a study, such as multiple elicitation formats). Flowers *et al.* (1997) used a computer-based bidding process and also found strong test-retest correlations averaging 0.796, which was comparable to other preference assessment methods. It should be noted, however, that this study had the benefit of a young and educated sample. O'Brien and Viramontes (1994) also used a bidding game and found the correlation to be lower at 0.66, comparing poorly against the standard gamble method, which produced a correlation coefficient of 0.82. Thompson *et al.* (1984) find the lowest correlation coefficient at 0.25. This is almost certainly due to the large time delay between the first and second iterations, at 1 year compared with only 4 weeks in the study by O'Brien and Viramontes (1994).

Test-retest reliability has been assessed in developing countries. The Pearson correlation coefficients ranged from 0.33 to 0.56. Cho *et al.* (2000) assess the test-retest reliability for malaria test kit in Myanmar. Retest was three weeks after test, and a bidding game format was used. The ICC ranged from 0.7 to 0.9. Dong *et al.* (2003) examine the test-retest reliability of the bidding game and dichotomous choice methods when used to elicit the WTP for community based health insurance in Burkina Faso. This study benefited from a larger sample size than previous studies, with 1284 members of the general population giving their WTP. The time interval between the test and retest was between four and five weeks. The Pearson correlation coefficients ranged from 0.593 to 0.675. Mean WTP is about 25% lower in the retest than in the test (statistically significant difference). A potential explanation for this is in terms of household food stocks. The first interview took place in January after the December harvest, whereas the second interview took place several weeks later when these stocks were at a lower level. The results suggest that the bidding game method is more reliable than the dichotomous choice method. Ultimately, the authors argue that the 'contingent valuation method as a whole appears to be a reliable technique for eliciting WTP' (p.2187).

More recently, Shiell and Hawe (2006) asked members of the Australian general public their WTP for a vaccine to reduce their likelihood of contracting an infection on holiday that would reduce their health to one of two EQ-5D health states. Telephone interviews were used with the bidding format of CV. In order to determine test-retest reliability interviews were repeated 1 week and 5 weeks after the initial interview. The Pearson's correlation coefficient between valuations at time 1 and time 2 ranged from 0.68 to 0.70. However, when comparing valuations at time 1 with valuations at time 3, the coefficients ranged from 0.36 to 0.54. There were also statistically significant differences between mean WTP at the different time points, with mean values increasing by 25% from time 1 to time 2 and time 3 (mean values were similar at time points 2 and 3). The authors conclude that the test-retest reliability of CV is 'acceptable, though not substantial' (p.180). There are also concerns that some respondents may have been giving consistent answers between time points due to their ability to remember their response in the previous time point. This would suggest the correlation coefficients are overestimates.

Smith (2004) assesses test-retest reliability by asking 47 members of the Australian general population their WTP to avoid being in a poor health state (as defined by the Assessment of Quality of Life Instrument), using an OE format. The valuation is repeated 4 weeks later. Depending on the particular state that is valued the Intra-class correlation coefficient ranges from 0.66 to 0.94. Smith argues the results indicate that 'WTP values can be reliable within the health context and from a general population sample' (p.36). An interesting finding is that the higher the WTP value is, the more reliable it is likely to be. This most likely due to an increasing imposition of the budget constraint. Qualitative feedback captured as part of the study is analysed in Smith (2007) and seems to suggest higher WTP values require more thought from respondents which makes them more stable, while at low levels of WTP values appear to be taken from a 'discretionary amount'.

In summary, the evidence on reliability of the CV method is mixed, with correlation coefficients ranging from 0.25 to 0.95, but most studies find the method to be reliable. One must be cautious when comparing across the studies because the time between the first and second valuation differs. Shiell and Hawe (2006) found significant differences in means which is perhaps more concerning for practical application of the results than individual variation.

3.2.2.3 Internal (Theoretical) Validity

The theoretical basis of the CV method, in the context of CBA, was outlined in the Introduction. Essentially, the method is strongly rooted in welfarist theory, as it estimates either compensating or equivalent variation (depending on whether WTP or WTA is elicited and whether a gain or a loss is being valued). Typically, two tests are performed to attempt to determine whether the results produced by the CV method comply with economic theory. Firstly, are WTP values positively associated with the size of the health gain being valued? Secondly, do WTP values increase with ability to pay?

To comply with economic theory WTP values should be sensitive to both the scale and scope of the good(s) being valued. When studying risk reductions in road traffic accidents Norinder *et al.* (2001) define insensitivity to scale as "a situation where respondents are unable to distinguish between different quantities of a particular good". They define scope insensitivity as "an effect where respondents are not able to separate their WTP for different outcomes". In other words, sensitivity to scale is an issue of the size/amount of the same service, while sensitivity to scope is an issue of the range of different services. However, the distinction between scope and scale insensitivity is not always made clear in the literature.

There is an extensive health economics literature studying the sensitivity in health of WTP values to scope and scale. Klose (1999) identify 15 references relevant to this issue (Acton, 1973; Thompson *et al.* 1982; Muller and Reutzel, 1984; Thompson *et al.* 1984; Tolley *et al.* 1984; Johannesson, Johansson, Kristrom, Borgquist, 1993; Johannesson, Johansson, Kristrom, Gerdtham, 1993; Neumann and Johannesson, 1994; Ryan *et al.* 1995; Ryan, 1996; Kartman, Johannesson, *et al.* 1996; Kartman, Stalhammar *et al.* 1996; Donaldson *et al.* 1997; Johannesson *et al.* 1997; Lee *et al.* 1997). Klose (1999) make no distinction between scope and scale, but instead refer only to 'scope'. All 15 references found that WTP values were positively related to the 'scope' of the good being valued. The references covered a wide range of outcomes including a reduction in the risk of becoming ill, a change in the severity of the illness and a change in the success rate of in-vitro fertilization. While the relationship between scope and WTP values was always positive, it was not always proportional and in some case showed declining marginal WTP. Declining marginal WTP is consistent with economic theory if there exists declining marginal utility of the good being valued. Determining whether insensitivity exceeds what would be expected as a result of diminishing marginal utility is a difficult judgement to make.

However, more recently studies have found evidence of scope/scale insensitivity. Olsen et al. (2004) varied the size of the health outcome both between respondents and within respondents. A total of 240 members of the Norwegian general public were asked to value heart operations for either 100 or 50 patients, cancer radiotherapy for either 300 or 150 patients or a helicopter ambulance that would save either 10 or 15 lives. A total of 300 members of the Portuguese general public were asked to value a reduction in the risk of having a heart attack for 150 patients by either 10%, 20% or 40%. The results showed no significant differences in WTP for different sized health effects. Even in the within sample comparisons, in which respondents were presented with two successive programmes which were similar in every respect except for the size of the outcomes, the majority of respondents stuck to the same WTP value for the programmes that had twice the effect. The authors conclude that 'these findings lend no support to the hypothesis based on the neo-classical theory of consumer behaviour that WTP should increase with the size of the good' (p.457). Chestnut et al. (1996), using a sample of 50 angina pectoris sufferers, found that mean WTP to avoid four angina attacks did not differ significantly from mean WTP to avoid eight attacks.

Smith (2001) and Yeung *et al.* (2003) test sensitivity to scale, where scale is defined as the level of health status (as a proxy for the quantity of 'health') possessed by an individual. Smith (2001) asks 47 members of the Australian general public their WTP to move from a poor health state to an improved health state, as described by the Assessment of Quality of Life (AQoL) instrument. An open ended question format is used and payment is out of pocket. Yeung *et al.* (2003) asked 142 Hong Kong residents their WTP to avoid either one day, three days or seven days of coughing, shortness of breath and congested throat with an open-ended question. Both studies found that sensitivity to scale of WTP responses appeared to decrease as the size of the health benefit being valued increased. In both studies the authors suggest this may be due to the increasing importance of the budget constraint as the value of the good increases relative to income. Smith (2005) further tests this theory through further analysis of the data presented in Smith (2001). Using regression analysis he provides further evidence in support of the importance of the budget constraint in determining sensitivity to scale.

Explanations for insensitivity to scope/scale, other than the role of the budget constraint, include the role of 'warm glow' (Kahneman and Knetsch, 1992). Utility may be gained from imagining that one contributes to good-cause programmes, where satiation may occur after some initial units of the good have been 'purchased', beyond which no more 'warm glow' is obtained. Thaler (1985) introduced the term 'mental accounting' for the anomaly that people seem to operate with separate budgets for various types of goods. When asked to express a valuation for a particular good, e.g. health care, the respondent taps into her 'mental' budget for health care, which has already been set. Therefore, if the mental budget for health care is low, the stated WTP reflects the upper level of this budget, rather than the trade-off between ones' income and the size of the described health improvement.

It seems evidence on the sensitivity to scope/scale of the CV method is mixed. Referring to the MVQ studies outlined in the previous chapter, Gyrd-Hansen (2003) acknowledges the likely presence of insensitivity to scope as changes in some dimensions of the EQ-5D did not influence WTP values. Baker *et al.*(2010) also found evidence of insensitivity to scope. WTP to avoid 12 months of headaches or stomach pains was only 2.2 and 2.3 times the WTP to avoid 3 months of headaches and stomach pains respectively. Prades *et al.* (2009) find insufficient sensitivity of WTP to both the duration of the health state and the size of the health gain.

Given diminishing marginal utility of income neo-classical theory would suggest that respondents with a higher income would have to give up more income to return them to the same level of utility following a health gain, therefore resulting in higher WTP values. There are numerous studies testing whether this is in fact the case. Klose (1999) identify 11 studies that show a significant positive correlation between income and WTP (Thompson *et al.* 1982; Thompson *et al.* 1984; Berwick and Weinstein, 1985; Appel *et al.* 1990; Johannesson *et al.* 1993; Neumann and Johannesson, 1994; Ryan, 1996; Ross *et al.* 1995; Chestnut *et al.* 1996; Kartman *et al.* 1996; Asenso-Okyere *et al.* 1997). Other studies have shown a positive relationship between social class (a proxy for income) and WTP (Donaldson *et al.* 1995; Miedzybrodzka *et al.* 1995; Donaldson *et al.* 1997). Klose (1999) was only able to identify three studies in which income did not have a significant influence on WTP (Johannesson *et al.* 1991; Johannesson *et al.* 1993; Olsen and Donaldson, 1993). This suggests that the CV method performs well against this particular test of theoretical validity.

More recently, Onqujekwe *et al.* (2008) tested the internal validity of a bidding game, dichotomous choice with follow up, and a novel haggling format when asking Nigerians to value insecticide treated mosquito nets. The amount of money spent on food is used as a proxy for income. The authors found that WTP was consistently found to be positively and significantly related to the expenditure on food. Grutters *et al* (2009) asked respondents their WTP for a hearing aid using a payment scale and open ended question. They found that there was a statistically significant but moderate relationship between WTP elicited through the open ended question and family income. WTP values elicited through the payment scale method were not correlated with family income.

Donaldson *et al.* (1995) also found that WTP was not associated with ability to pay (as predicted by theory) for the open ended questions, but it was for the payment card method.

Therefore, the authors conclude that the payment card method is more valid than the open ended approach.

This section has outlined studies performing one of two tests of internal validity; sensitivity to scale and correlation with income. The results on sensitivity to scale are mixed. Diminishing marginal WTP as the size of the good being valued increases complies with economic theory, but in some case insensitivity seems to exceed what would be expected as a result of diminishing marginal utility. It is likely that the increasing importance of the budget constraint plays an important role in limiting the sensitivity of responses. Regarding correlation with income, it seems that the CV method performs well against this particular test.

3.2.2.4 Construct Validity

3.2.2.4.i Convergent Validity

As outlined earlier in the introduction CV questions can be asked in four different fashions: Open-ended, Payment Scale/Payment Cards, Closed-ended/Discrete choice, or bidding games. Each approach has its own strengths and weaknesses (Drummond et al. 2005). Open Ended questions pose a difficult cognitive task for most respondents because people are typically not used to thinking about the maximum they would pay for something. Furthermore, the method can elicit large numbers of protest responses i.e. the respondents refuse to give a value (O'Brien and Gafni, 1996). Bidding games may improve the precision of estimates, but at the expense of starting point bias i.e. the maximum WTP value a person gives is affected by the first value they are presented with. The Closed Ended/Discrete Choice Method is reliant upon econometric techniques to identify the quantitative relationship between the proportion of persons accepting or rejecting the bid at different levels and hence estimate a WTP value. This approach requires large sample sizes and the identification of the relevant range in which to sample bids. Weaknesses specific to a particular method, including starting point bias in bidding games and range bias in payment cards, are discussed in separate sections. This section focuses on studies directly comparing different elicitation methods.

There is no obvious conceptual reason why the results of the different methods should vary (O'Brien and Gafni, 1996). Indeed, the measures are all intended to measure the same object. However, some evidence suggests there may be systematic differences between the methods. This section compares the methods in terms of the results produced.

(a) Open Ended vs Payment Card

Klose (1999) identifies three studies that elicit WTP values through both the open-ended and payment card methods (Miedzybrodzka *et al.* 1994; Miedzybrodzka *et al.* 1995; Donaldson *et al.* 1995). In all three studies the payment card method produced higher WTP values. In two of the studies payment cards produced mean WTP values 25-30% higher than the Open Ended approach (Miedzybrodzka *et al.* 1994; Miedzybrodzka *et al.* 1995), while the other study found card values to be only 10% higher (Donaldson *et al.* 1995).

In her PhD thesis Emma Frew elicited WTP for two forms of colorectal cancer screening, using all four WTP elicitation formats (Frew 2003). The main part of the study, described in Frew *et al.* (2001), distributed self-complete questionnaires to patients (who did not necessarily have any experience of colorectal cancer screening) at English GP practices. WTP was elicited through OE questions and through Payment Scale. A total of 2214 returned questionnaires were useful for further analysis. A paper by Whynes *et al.* (2003) directly compares the two WTP elicitation formats. Through logistic regression analysis they find the payment scale format to lead to higher mean WTP values. The explanation offered is of a framing effect. Since the payment scale offered a range which extended to 20 times the median valuation this may have 'dragged up' individual valuations. In contrast to the above studies Grutters *et al.* (2009) asked hearing aid users their WTP for a hearing aid using both the payment card and open ended approaches and found no significant differences in the values produced.

(b) Discrete Choice vs the others

Klose (1999) identifies two studies comparing the OE method with the DC method (Johannesson *et al.* 1991; Donaldson *et al.* 1997). Johannesson *et al.* (1991) judged the DC approach to be problematic since the answers were strongly influenced by the true costs stated within the questionnaire. They found that the DC approach led to estimates that were 12 times greater than the open ended method. Donaldson *et al.* (1997) also found the discrete choice method to be problematic and note their concern that many respondents estimate cost and state this as their WTP.

Frew et al. (2003) replicate their earlier study (Frew et al. 2001) by asking 354 members of the UK general public their WTP for two different methods of screening for colorectal cancer (also elicited through self-complete questionnaires). The more recent study differs in that the WTP format is DC (with analysis through logistic regression). They find that the Discrete Choice method produces significantly higher results than the other two approaches, by a magnitude of approximately 20 times when compared with the Payment Scale format and 33 times when compared with the OE format.

Evidence from a study by Ryan et al. (2004) supports the findings of Frew *et al.* (2003). They compared Payment Card questions with Discrete Choice questions through interviews with 578 members of the Scottish general population. WTP was elicited for three healthcare interventions: an expansion in cancer treatments, an expansion in heart operations, and the introduction of a helicopter ambulance service. The payment card questions elicited mean WTP values ranging from £40 to £54, while the discrete choice questions elicited values ranging from £122 to £356 (depending on which treatment was being valued, the upper limit used in the discrete choice questions and the treatment of "do not knows").

Kartman *et al.* (1996) compared the discrete choice approach with the bidding game approach by using the first answer of the bidding game as the discrete choice answer. The patterns of estimated WTP were similar leading the authors to conclude that the discrete choice method should be preferred because the risks of starting point bias in bidding games were to great.

(c) Bidding Game vs The Others

Frew *et al.* (2004) also make comparison with their earlier study (Frew *et al.* 2001). They conducted bidding game format WTP interviews with 106 members of the UK general population, to obtain their preferences for the two methods of colorectal cancer screening. They found that the bidding game format produced WTP estimates approximately seven times larger than the payment scale and open-ended formats used in Frew *et al.* (2001). The predominant explanation proposed by the authors is interviewer bias. The Frew *et al.* (2004) study was conducted by face to face interviews, while the Frew *et al.* (2001) study was conducted through postal questionnaires. Interviewer bias occurs when respondents "shape their answers in a way that they think will either please the interviewer or will increase their status in the interviewer's eyes" (Mitchell and Carson, 1989). Respondents might be unwilling to offer low values to an interviewer evidently positively disposed towards the service being valued. This explanation suggests that at least some of the cause of the disparity in results may be endogenous to the WTP format. The Frew *et al.* (2004) study also found evidence of starting point bias in the bidding game format. For one of the two cancer screening methods, when the starting bid was £10 the mean WTP was £341, but when the starting point was £200 the mean WTP was £607. The WTP values elicited through the bidding game format in Frew *et al.* (2004) were lower than those elicited through discrete choice in Frew *et al.* (2003).

(d) Comparing all four methods

Heinzen and Bridges (2008) use all four methods to value a pneumococcal vaccine in Bangladesh. They found that WTP values varied significantly across the methods, with average estimates varying between $2.34 and $18. The OE method produced the highest results while the payment cards produced the lowest results (the discrete choice method was not used to elicit a WTP value).

(e) Convergence between CV and other Preference Elicitation Methods

Some studies have considered the correlation between WTP and utility elicited through VAS, SG or TTO. Zethraeus *et al.* (1997) found that the QALY-weight gain of a hormone replacement therapy for women with severe symptoms was about two times higher than for women with mild symptoms (1.9 times higher for VAS and 2.3 times higher for SG). The WTP was 2.1 times higher in the severe symptom group, showing a correlation between the measures of utility and WTP at the aggregate level. Coley *et al.* (1996) found that VAS, SG and WTP were all consistent when used for ordering scenarios of pneumonia therapy. However, O'Brien and Viramontes (1994) found that WTP could not differentiate between the severity of chronic lung disease in contrast to utility measurements by either VAS or SG methods.

(f) Summary of Convergent Validity

It seems that the results of CV studies vary widely depending on the elicitation format used. Typically OE questions seem to give considerably lower results than the other methods. The MVQ studies reviewed in the previous chapter employed differing elicitation methods. The results suggest that making comparisons between studies using differing methods is difficult. Therefore, knowing which MVQ estimate is the most appropriate for potentially informing policy decisions is challenging.

The need for standardisation has been recognised has been recognised in the environmental field. The report of the US National Oceanic and Atmospheric Administration (NOAA) expert panel has been particularly influential in this respect (Arrow et al. 1993). The panel advocated the use of the discrete choice format in preference to Open-Ended. This was on the basis that the Closed Ended format is more realistic and respondents would have no incentive to behave strategically. Since the recommendations of the panel the discrete choice format has also been recommended as the standard method for use in health economics (Johannesson et al. 1996). However, this has not been widely accepted (see Smith, 2000), not least because the recommendations of the NOAA panel were with respect to the environmental arena and the health care field is quite different.

3.2.2.4.ii Discriminant Validity

Differences between the elicitation methods, reviewed in the previous section, may be driven by potential sources of bias associated with each of the elicitation methods. Klose (1999) posits that biases of the CV method can be based on 'response effects', on 'sampling' and on 'inference'. The response effect bias is subdivided into 'incentives to misrepresent responses', 'implied values cues' and 'scenario misspecifications'. Incentives to misrepresent responses may occur, when respondents state a WTP amount which differs from their true WTP amount because of strategic reasons, or as an attempt to comply with expectations of sponsors or interviewers (Strategic Bias). Implied value cues occur when the contingent market scenario provides or is interpreted to provide information of a correct value. Examples are stating point bias in bidding games and range bias of payment cards. The dichotomous choice method may also lead to bias when respondents tend to answer 'yes' incorrectly to the bids ('yeah saying'). Scenario misspecification biases occur when the scenario is not understood by respondents as intended. An example is question order bias which is said to occur when the order of the questions influences the stated WTP amounts. Examples of each of these biases and studies testing them are reviewed below.

(a) Starting Point Bias

Starting Point Bias may arise in a Bidding Game format if the final value that a respondent gives is influenced by the first value they are faced with. Klose (1999) identifies five references relevant to starting point bias (O'Brien and Viramontes 1994; Kartman *et al.* 1996; Stalhammar, 1996; Chestnut *et al.* 1996; Phillips *et al.* 1997). All of these studies found evidence of starting point bias, with one exception (O'Brien and Viramontes, 1994). In the study by Stalhammar (1996) a total of 105 Swedish users of anti-secretory drugs were asked how much they would be willing to pay for a more convenient drug that could be taken with meals, through one-off out of pocket payment. Half of the sample received the high bid first (1000 SEK), and the other half received the low bid first (20 SEK). On average, those who received the high bid first were prepared to pay 289 SEK, compared with 70 SEK for those who received the low bid first.

Searching for articles that had cited Stalhammar (1996) identified two more recent articles testing for starting point bias in CV (Bhatia, 2005; McNamee *et al.* 2010). Both studies found evidence of starting point bias, as did the study by Frew *et al.* (2004). In response to the study above that did not find evidence of starting point bias (O'Brien and Viramontes, 1994), McNamee *et al.* (2010) argue that it is blighted by a small sample size (n=102), which may explain the failure to reject the null hypothesis of no starting point bias.

Starting point bias can potentially be overcome by the use of an alternative method of administering the CV exercise, such as Open-Ended. However, these alternative methods are also plagued with difficulties, so the researcher must weigh up which approach he feels will provide the most accurate results. Furthermore, Stalhammer (1996) has suggested that if starting point bias is present this indicates a more serious problem that cannot be overcome by using an alternative method. It may suggest that respondents' preferences are fundamentally unstable and hence pose questions over the whole CV approach.

It is hard to know the likely effect of any starting point bias in the MVQ studies using a bidding game format, and the authors do not hypothesise over any likely effect (Blumenschein and Johannesson 1998; King *et al.* 2005; Lieu *et al.* 2009; Shiroiwa *et al.* 2010; Zhao *et al.* 2011). In the study by King *et al.* (2005) the initial bid was $1, the second bid was equivalent to the respondents' monthly household income, and the maximum bid was 10 times the respondent's annual household income. If a respondent's real value was considerably higher than the first bid, then this may have exerted a framing effect which could lead them to revise their valuation downwards. In the other two studies the bid levels are not stated.

(b) Range bias

Range bias is similar to starting point bias, and can arise in the payment card format if the value a respondent gives is influenced by then range of values presented on the payment card. Evidence on the existence of range bias is limited and inconclusive. Klose (1999) identify two studies relevant to range bias. The first of these (Neumann & Johannesson 1994) asks 231 US potential child-bearers their willingness to pay for IVF treatment. In one section of the survey respondents were asked if, in the event that they were infertile, they would be prepared to pay a number of stated amounts: $100, $5,000, $10,000, $25,000, $50,000, $100,000 or $200,000 for IVF with different chances of success: 10%, 25%, 50%, and 100%. However, to test for range bias monetary values that were twice as large were presented to 20% of the respondents. On average the WTP values in the sample given the double amounts were about 30% higher. The other study identified by Klose (1999), (Johannesson et al. 1991), asked 400 patients their WTP for treatment for hypertension using two different CV questions with slightly different ranges on the payments cards in each version. The study found that the payment cards with a broader range elicited higher WTP values, but these findings were not significant. Ultimately, Klose (1999) concludes that there are 'only hints of range bias' (p.105).

More recently, Whynes et al. (2004) asked 1401 members of the UK general population their WTP for colorectal cancer screening. Two different payment cards were used, and respondents were randomised to receive one of the two. The broad payment card extended to £1000, while the shorter version only extended to £100. The results showed that the long-scale instrument produced a mean WTP more than 30% higher (statistically significant) than the short-scale instrument. The authors conclude as follows: "We believe our findings to be strongly supportive of the existence of range bias in payment card instruments" (p.189).

In response to the problem of range bias some recent CV studies have used an alternative elicitation method that involves a randomised card sorting (RCS) procedure (Carthy *et al.* 1999; Chilton *et al.* 2004; Smith, 2006). Whereas the payment card approach lists all values on one card, the RCS approach has separate cards for each individual value. The cards are shuffled in the presence of respondents and then drawn one at a time in a random order. Respondents are asked to sort the cards into one of three categories: amounts they are sure they would pay, amounts they are sure they would not pay, and amounts they are unsure about. It is argued that this series of 'take it or leave it' choice simplifies the valuation task and hence increases the respondents engagement with the exercise. Furthermore, by avoiding explicit presentation of the full range of values to the respondents, the RCS may hope to attenuate the range bias described above. It is hoped that by shuffling the cards in front of the respondents starting point bias will be avoided since it will indicate to respondents that the first value gives no indication of the 'correct' answer (Covey *et al.* 2002).

Smith (2006) tested the RCS approach against two payment card approaches, one which presented bids in a low to high ordering, and another which presented bids in a high to low ordering. A sample of 314 members of the Australian general public were asked their WTP for a number of health states described by the Assessment of Quality of Life instrument (Smith, 2001), with payment being out of pocket. The high to low version produced significantly higher values than the low to high values and the RCS version. Validity tests were performed by asking respondents to distinguish between values that they were 'sure they would pay'. Values they were 'sure they would not pay', values they were 'sure they would not pay' and the value that they felt 'most closely approximates their maximum WTP'. In the RCS format significantly fewer respondents had values they were 'unsure' they would pay, and fewer instances where such values were higher than their expressed maximum WTP. The author therefore argues that values obtained from the RCS approach are more likely to reflect 'true' WTP. In light of this, Smith argues that the high to low version is likely to be subject to starting point bias which increases valuations.

In contrast to the findings of Smith (2006), Covey *et al.* (2007) find that the RCS procedure is no less vulnerable to range bias than the payment card method when applied to health risk reductions and death rates. A sample of 240 members of the UK general public were asked their WTP for safety features which might be added to their car, through either the payment card or RCS approach. Within each of these approaches respondents were randomly allocated to be presented with amounts, either on one card, or on multiple cards, ranging from £0-£500 or £0-£1500. The effect of range bias was higher in the RCS approach than in the payment card approach. The ratio between the high and low range mean responses was 1.18 for the payment card approach, compared to 2.01 for the RCS approach.

It is clear that the MVQ estimates based on the payment card format may have been different, had a different range of values been used. Therefore, it is essential that a strong justification for the range of values used is presented. However, of the MVQ based studies employing the payment card format only one attempts to provide such a justification (Cunningham and Hunt, 2000). This justification is weak: the authors argue that the range was acceptable to respondents (although how this is determined is not clear), and they point out that the actual cost of the procedure lied within the specified range. One must be cautious even when comparing MVQ estimates based on the same WTP elicitation format.

(c) Strategic Bias

Free-riding may occur in WTP in two directions. Firstly, if respondents think they will actually have to pay the amount they reveal they may underbid. Alternatively, if respondents do not believe they will actually have to pay their stated WTP amount, but they want to influence the provision of the good in question, we might expect them to overbid. One approach to overcome the first type of free-riding behaviour is to make the scenario as hypothetical as possible. However, this will obviously increase the likelihood of encountering the second type, as well as the other problems noted above that are inherent when the scenario becomes too hypothetical.

Klose (1999) identified only one reference that tests for strategic bias (Phillips *et al.* 1997) and the study finds no evidence of its existence. Johannesson and Jonsson (1991) acknowledge that the hypothetical character of CV may increase variance, but no evidence exists that there is a bias in any one direction. Therefore they did not include 'hypothetical bias' in their typology of potential biases.

More recently, Hackl and Pruckner (2005) test for free-riding by asking 2536 members of the Austrian public their WTP for the provision of health-related Red Cross services. Version 1 of the questionnaire asked people to make a donation, and stated that only those who had donated could benefit from Red Cross services. In Version 3 people were told that anyone could benefit from the services regardless of whether they had made a donation (Version 2 was used to assess warm glow). Each respondent completed all three versions (any randomization is not made clear). If respondents were exhibiting free-riding behaviour they would have an incentive to give a lower WTP value in Version 3 compared to Version 1. However, the authors found little evidence of free-riding behaviour. Their results were robust and not suggestive of strategic behaviour.

(d) Question Order Bias

Order effects may occur within a study wishing to elicit WTP values for multiple programmes. Stewart et al. (2002) outline three reasons why a study may seek to do this. Firstly, the NOAA guidelines (Arrow et al. 1993) recommend that respondents be reminded of substitutes when reporting WTP values. Secondly, if the results are to be used to inform resource allocation decisions then valuations are required for multiple programmes. Finally, if faced with only one programme respondents may feel that this is representative of the wider health sector and hence overestimate their valuation without any notion of opportunity cost. However, order effects arise if the valuation given to a specific programme is influenced by the order in which the programmes are presented to the respondent.

Klose *et al.* (1999) identify two studies relevant to ordering bias (Olsen and Donaldson, 1993; Kartman *et al.* 1996). The first of these found evidence suggestive of ordering bias (Olsen and Donaldson, 1993), but as it is not designed to specifically test for this bias, drawing firm conclusions is difficult. The second study specifically tested for ordering bias and was unable to find evidence for its existence (Kartman *et al.* 1996). More recently, Stewart et al. (2002) presented 473 members of the Irish population with three completely different programmes: an increase in pain-relieving treatments for cancer patients (C): an increase in heart operations (H), and an increase in community care services (CC). The first of two sub-samples valued the programmes in the order C, H, CC, while the other sub-sample faced the programmes in the reverse order: CC, H, C. Respondents were first asked to rank the different programmes. They were then asked if they would be willing to contribute anything through extra taxation for the first programme. Next they were asked if they would be willing to pay through a voluntary contribution. Finally, they were asked the maximum their household would be willing to pay each year for the expansion (out of pocket). This process was then repeated for the other two programmes. The study finds ordering effects in the ranking of the programmes, in the proportion of zero values reported and in the WTP for one of the three programmes (community care). The primary explanation for these effects offered by the authors is fading glow: "the first programme in any sequence captures much of the utility associated with giving" (p.585).

A number of explanations have been proposed for ordering effects. Carson *et al.* (1998) found that the value of a public good tends to fall when it is valued further down in the sequence of goods. They argued this effect can be predicted by economic theory due to the substitution possibilities between the goods and the reduction in disposable income that occurs with the purchase of each new good. Stewart *et al.* (2002) argue ordering effects exist because of a fading glow effect. They argued that respondents tend to receive 'warm glow' or moral satisfaction from contributing to a publicly financed good, and that the first program in the sequence is likely to capture the bulk of this moral satisfaction. Therefore, respondents are prepared to pay more for the first program than for subsequent programs.

Two of the MVQ studies are likely to be susceptible to order effects as they ask the WTP for a number of different outcomes, and do not vary the order of the questions (Mason and Donaldson, 2007; Prades *et al.* 2009) . Since these studies do not specifically test for ordering effects it is impossible to predict their likely influence. It is important that studies administering a series of WTP questions for different outcomes randomise the order of the questions to minimise ordering effects at the aggregate level.

(e) 'Yeah saying' bias

'Yeah saying' bias occurs if respondents tend to answer yes to a dichotomous choice question even though their true WTP is smaller than the amount presented in the question (Kanninen, 1995). One possible explanation for such behaviour is that respondents use the bid as an exogenous information about the value of the health technology. Hammerschmidt *et al.* (2003) state that only one study in the health care literature had examined 'yeah saying' behaviour. In a study by Chestnut *et al.* (1996) a series of closed-ended questions were followed by a direct question. Ten percent of the respondents gave lower responses to the direct question than the highest amount they had said 'yes' to in the closed-ended question. The authors believe this to be possible evidence of 'yeah-saying' behaviour.

Hammerschmidt *et al.* (2003) test for both 'yeah-saying' bias and 'nay-saying' bias, (respondents tend to answer no to a dichotomous choice question even though their true WTP is higher than the amount presented in the question), by responses to dichotomous choice questions with those from payment card questions. Patients with Type 2 diabetes are asked their WTP for a reduction in the risk of diabetic complications firstly through payment card, and then through dichotomous choice. The bids in the dichotomous choice questions were based on the distribution of WTPs which the patients were definitely willing to pay according to their answers on the payment cards. If a respondent stated a willingness to pay a given amount in the payment card exercise, and then refused to pay this amount in the DC questions this was regarded as an 'unexpected no', and visa versa. Of 102 'yes' responses to the DC questions 31 were deemed to be unexpected. Of 72 'no' responses 16 were unexpected. No-answering behaviour was more pronounced at low bids, while yes-answering behaviour predominated at high bids. The study design allows only tentative explanations for the findings. Nay-saying might theoretically be explained by conservatism or when preferences for money improvements dominate. However, it could be that range bias in the payment card questions may have caused an upward bias in these questions, leading to unexpected 'no' responses in the DC questions. Yeah-saying may have arisen when health improvements dominate over money-improvements.

(f) Summary of Discriminant Validity

Earlier it was shown that the different elicitation methods can produced widely differing results in CV studies. This section has shown potential sources of bias that may drive these differences and highlights the careful consideration researchers must give to planning a CV study in order to minimize bias. There is evidence of starting point bias in the bidding game format, range bias in the payment card format, and yeah saying bias in the DC format. No evidence of strategic bias was found but question order bias represents another potential concern.

3.2.2.5 Other Conceptual Aspects of Study Design

As well as elicitation method, addressed earlier, there are other conceptual aspects of the study design that may influence the results produced by a CV study. In particular, evidence suggests that whether WTP or WTA is used and the choice of payment vehicle that is specified may influence results.

(a) WTP or WTA

As outlined in the introduction, CV can elicit either the WTP or WTA. The choice of method will be determined by the direction of measurement and whether the researcher wishes to elicit the compensating variation or equivalent variation. If the introduction of a technology induces a utility gain, the compensation variation is a willingness to pay, the equivalent variation is a willingness to accept. Both are monetary measures of the same utility gain. However, some evidence suggests that WTP and WTA do not coincide. While there is extensive evidence in the environmental field, it is somewhat limited in the healthcare context. The review by Klose (1999) identifies only one study that directly compares WTP and WTA (Donaldson *et al.* 1997). In this study WTP was measured as compensating variation and WTA as equivalent variation. There were no significant differences in valuations derived from the two methods. More recently, differences between WTP and

WTA for visits by a family physician have been studied (Fernandez et al. 2010). A total of 451 Spanish subjects at six health centres were asked their WTP followed by their WTA for visits by a family physician. Payment cards were used and the payment vehicle was one-off out of pocket payment. They found that, on average, WTA values were 3.3 times higher than WTP values. They also found the ratio increases with age and decreases with income. Elsewhere, in the healthcare context, O'Brien et al. (1998) evaluated a new drug to treat cancer and found a WTA/WTP ratio of approximately 2.

---Reasons for the disparity between WTP and WTA

One explanation is the existence of an 'income effect'. WTP is constrained by ability to pay, while WTA has no such constraint. Therefore, it is possible that payment capacity in WTP is reached before satisfaction with the compensation in WTA is reached. Another theory, is Kahneman and Tversky's (1979) psychological theory of an endowment effect, such that the loss of utility associated with the loss of something weights heavier than the utility associated with gaining the same programme benefit. Hanemann (1991) has argued that a lack of substitute commodities for a removed programme will inflate WTA.

---Implications of this disparity

Since WTP is used almost exclusively in health care CV studies, one may feel that the discrepancy between WTA and WTP is of little importance. Indeed, Smith (2003) reviewed 111 CV studies and found only one that had applied WTA (Lindholm, Rosen and Hellsten, 1994). However, O'Brien et al. (2002) argue that the discrepancy between WTA and WTP is evidence that members of the general public value an intervention differently depending on whether it is being introduced or phased out (similar to the theory of Kahneman and Tversky, 1979). In other words, the 'selling price' of an intervention is different to the 'buying price'. Not only does this suggest that cost benefit analysts need to be careful to adopt the right approach, it also has implications for cost-effectiveness analyses using the QALY as the measure of benefit. O'Brien et al. (2002) argue that the disparity suggests a kink in the cost-effectiveness threshold, so that the slope is different in the NE quadrant of the CE plane (where the cost and effects increase and hence benefits would be measured through WTP), compared to the SW quadrant (where the costs and effects fall and hence the negative benefits are measured through WTA).

The MVQ studies in the previous chapter all apply WTP. If they had instead applied WTA the likelihood is that the MVQ estimates would be considerably higher. A potential implication of the argument of O'Brien *et al.* (2002) is that the MVQ estimates can only be applied to programmes that are being introduced, not programmes that are being phased out.

(b) Payment Vehicle

Payment vehicle refers to the way in which the hypothetical payment in the CV market is to be made e.g. through taxation or out of pocket. There is limited literature studying the impact payment vehicle has on results of contingent valuation studies. O'Brien and Viramontes (1994) compared out of pocket payment WTP for IVF with WTP in taxes for a public IVF program. Although the public perspective should cover additional aspects such as altruism in the evaluation, the personal WTP was higher. This might also indicate a negative attitude against taxes as payment vehicles.

Watson and Ryan (2004) asked 1400 members of the UK general public their WTP for air ambulances through computer assisted telephone interviews using a discrete choice format. Two payment vehicles were used, taxation and donation, with respondents randomly allocated to one of the two vehicles. The first finding was that more respondents were prepared to make a contribution through taxation (61%) than were prepared to make a charitable donation (54%). The second finding was that respondents facing the taxation questions gave higher WTP values (i.e. they were more likely to state 'yes' to the willingness to pay amounts).

The foremost explanation for payment vehicle bias is strategic behaviour as outlined earlier (although the evidence found did not support the hypothesis of strategic behavior). There may be an incentive to exaggerate WTP when payment is not compulsory. Conversely there may be an incentive to state too low WTP when payment is individual and compulsory. However, such an explanation does not explain the findings of Watson and Ryan (2004), that voluntary donation leads to lower WTP values. In this case individual may not choose to make a contribution (or make a low one) if they feel others will not make a similar contribution. Taxation may elicit greater WTP because taxation is progressive and will involve making payments in small increments over a long period of time. This is in contrast to out of pocket donations which will require a one-off payment. Furthermore, one-off payments are more likely to be constrained by ability to pay than contributions through taxation.

The payment vehicle adopted must be appropriate for the jurisdiction in which the questions are being asked, to make the contingent market as realistic as possible, and to avoid the potential sources of bias outlined above. For example in health, taxation may be appropriate for the UK, but insurance premiums would be more relevant for the US. Furthermore, as noted by Drummond et al. (2005), different payment vehicles will be appropriate for the valuation of different products. So, insurance may be suitable for high technology items, while out of pocket payments (e.g. co-payments) may be most relevant for consumer based products such as pharmaceuticals.

Smith (2003) reviewed 111 health care CV studies and noted the payment vehicle used. The following categories were used: out of pocket payment; taxation; private insurance; voluntary donation; combinations of these; not available. In the vast majority of the studies (82%) out of pocket payment was the vehicle of choice. Of the remaining studies 10 fell into the 'not available' category, five used taxation and five used a combination of methods. Cross tabulation found that choice of payment vehicle was not systematically influenced by the country in which the study was being conducted. Tax payments were not used in the UK in any of the studies, and insurance was not used (apart from in combination with out of pocket payments) in any US studies. Smith (2003) comments, "There seems to be some confusion in the conduct of studies" (p.615).

Of the existing MVQ studies all but one use out of pocket payment, which is consistent with the existing healthcare CV studies reviewed by Smith (2003). The effect that this has on values is unclear. Ability to pay may act as a greater constraint when one-off payments are used, leading to lower estimates than if alternative methods such as taxation were used. As outlined above the effect of the payment vehicle may differ depending on the jurisdiction in which it is applied. For example, Byrne *et al.* (2005) use a one-off out of pocket payment in the US, while Mason and Donaldson (2007) use the same approach in the UK. Given the presence of a national health service in the UK this approach may be particularly inappropriate. Given that very few health care systems around the world now require patients to pay out of pocket for their treatment at the point of use, beyond small co-payments for pharmaceuticals (especially in the developed countries where the MVQ studies have been performed), the contingent markets in the existing MVQ studies have perhaps not been appropriately specified.

This section has shown that the payment vehicle specified in a CV study is an important consideration. The potential impact of the budget constraint will be influenced by the choice of vehicle, and if the results are to reflect real life behavior (criterion validity, assessed in the next section) the vehicle must be realistic to the jurisdiction in which the questions are asked.

3.2.2.6 Criterion Validity

The strongest test of validity may be seen as the extent to which behaviour stated in surveys is replicated in the real world. There are limited studies testing this in health economics, partly due to the lack of a market for most healthcare interventions. Klose (1999) identifies two studies that allow some assessment of criterion validity (Walraven, 1996; Anderson *et al.* 1997). Walraven (1996) asked fee paying hospital inpatients and outpatients how much they would be willing to pay for the services they were receiving. The majority of respondents stated WTP values that were less than the amount they were actually paying. This suggests that the fees are considered too high which deters some potential users leading to an even higher average cost for the actual users. Fees may be too high, but when faced with need, patients pay more than they would normally be willing to. Alternatively, the design of this study may have been susceptible to strategic behaviour with patients seeking to influence future prices. Anderson *et al.* (1996) asked respondents their WTP to shorten their waiting time for elective surgery. Between 12-25% of respondents indicated that they were willing to pay the market rate (by going private) but only 1.7% had actually chosen to do so. This suggests that respondents exaggerate their true WTP.

Bhatia and Fox-Rushby (2002) assessed criterion validity by asking 300 households in India their WTP for treated mosquito nets with a bidding game format. One month later the households were revisited and asked whether they would be prepared to buy a treated mosquito net for the modal value from the previous survey. The results showed that at the aggregate level their was no discrepancy between the hypothetical and actual values. However, at the individual level around 35% of respondents did not act in accordance with their hypothetical preferences. A similar approach was taken by Onwujekwe *et al.* (2005) who asked members of the Nigerian general population their WTP for insecticide treated bed nets through bidding game, dichotomous choice, and a novel structured haggling technique. One month after this initial survey respondents were given the opportunity to buy the nets at a fixed price (less than the market selling price). In total 21 out of 71 respondents, who initially stated a WTP greater than the asking price in the second survey, would not buy a bed net. Statistical analysis showed that the key explanations for these divergences were the attitude of community leaders and the external information the respondents received, both of which increased the initial WTP value.

Blumenschein *et al.* (2001) conducted a field experiment comparing hypothetical and real purchase decisions for a pharmacist providing asthma management programme among 172 subjects with asthma. Subjects in the hypothetical group were asked their WTP through a DC question using one of three prices, while subjects in the actual group were given the opportunity to purchase the service at one of the three prices. In the hypothetical group 38% of subjects said that they would purchase the good at the stated price, but only 12% of subjects in the real group purchased the good. There were statistically significant differences between the hypothetical and actual groups for two of the three prices. However, subjects in the hypothetical group who stated they would pay the given amount were asked whether they were 'probably sure' or 'definitely sure' of their decision. The overall proportion of 'definitely sure' yes responses was 14%, which was very close to the 12% yes responses in the real group. When only the 'definitely sure' yes responses were included there were not statistically significant differences between the groups for any of the prices. However, the authors acknowledge that this may be coincidental and the finding does not validate the particular calibration method.

3.2.2.7 Summary of findings on the CV method

Extensive evidence has been presented on the performance of the CV method. Results on feasibility are mixed with completion rates ranging from 69% to 99%. Non-completions are typically due to protest responses. The results on reliability are also mixed. The differences between studies performing test-retest reliability analysis make it difficult to draw firm conclusions. In terms of validity, the method seems to correlate well with income which is consistent with theory. Evidence on insensitivity to scale is mixed, with the budget constraint revealing itself as a potential source of any insensitivity. One of the key findings is that results of CV studies seem to vary greatly depending on the elicitation format used. This variance is potentially driven by a number of sources of biases that the different formats are susceptible to.

3.3 Time Trade-Off

As mentioned in the introduction, the TTO method requires participants to choose between two scenarios and make and in doing so make a trade-off between quality of life and length of life (Torrance *et al.* 1972). The method was developed to generate values similar to the SG method, while avoiding the need for the use of complex probabilities (Buckingham and Devlin, 2006). This section assesses the feasibility, reliability and validity of the method.

3.3.1 Search Strategy

As with the previous review on the CV method, the aim of this review is to consider the feasibility, reliability and validity of the TTO method. A review of reviews was considered the most appropriate method of capturing the relevant literature. These could be systematic reviews or theoretical/methodological discussions of the method in the context of health. The review by Green *et al.* (2000) was used as a starting point. Note, this review covers up to November 1997. A further three reviews were cited by Green *et al.* (2000) (Torrance, 1986; Torrance, 1987; Froberg and Kane, 1989). In order to identify more recent work a pearl growing approach was used with the most recent review, Green *et al.* (2000) used as the starting point. A total of 125 articles were identified (as of November 2011), and a total of seven were deemed relevant for inclusion. An article was included if it was designed to test a specific facet of feasibility, reliability or validity. Studying the reference lists of these articles led to the identification of a further two review articles (Nord, 1992; Dolan, 1998). Pearl growing from these other review articles identified a further eleven articles. Finally, although Dolan *et al.* (1996a) is not a review (it is an empirical study comparing different variants of TTO and SG), since it is a highly referenced article a pearl growing approach was used with this article as the starting point. This led to the identification of a further ten articles, three of which were reviews (Arnesen and Norheim, 2003; Arnesen and Trommald, 2004; Arnesen and Trommald, 2005). At this point it was deemed that saturation point had been reached.

In summary a total of nine review articles and a further 27 articles since 2000 were identified that address the feasibility, reliability and validity of TTO.

3.3.2 Findings

3.3.2.1 Feasibility

In their review, Froberg and Kane (1989), identify only one study on the feasibility of the TTO method, which finds TTO to be easier for respondents than SG (Torrance, 1976). Torrance (1976) measures the feasibility of VAS, TTO and SG by their acceptability to the subjects, their ease of use for the interviewers and their cost. The results showed all three methods to be acceptable to respondents as only 2% of all interviews were broken off by respondents. However, the TTO method was found to be the easiest to complete, based on informal feedback from three of the samples, and more formal measurements introduced with a fourth sample. The interviewers found all three techniques easy to learn and administer and had no preference for one over the other. Despite this, the authors note that without colour coded cards and a probability wheel the SG method would have been almost impossible. Furthermore, the SG method was only used on the most educated of the four samples. Torrance acknowledges that the TTO and SG methods are both time consuming and expensive, typically requiring trained interviewers. In contrast, the VAS method is more straightforward and could be administered as a postal questionnaire, perhaps as an add-on to existing household surveys.

Green *et al.* (2000) identify eleven studies that support the feasibility (they use the terminology 'practicality') of TTO (Detsky *et al.* 1986; Krumins *et al.* 1988; Fryback *et al.* 1993; Reed *et al.* 1993; Ashby *et al.* 1994; Glaziou *et al.* 1994; Patrick *et al.* 1994; van der Donk *et al.* 1995; Dolan *et al.* 1996a; Johnson *et al.* 1996; Kreibich *et al.* 1996). These studies demonstrate high completion and response rates, and three of the studies have shown the TTO method to outperform the SG method by small margins (Red *et al.* 1993; Dolan *et al.* 1996a; van der Donk *et al.* 1996). Fryback *et al.* (1993) perform a large scale population survey in the US with 1,356 respondents who are asked to value their own health through TTO. They also achieve response rates of over 95%. Glasziou *et al.* (1993) has found TTO to be feasible in self-complete format. The achieved completion rates of 91% in a sample of 714 patients 6 months after myocardial infarction. SG has also been shown to be feasible in self-complete format (Dolan *et al.* 1996a).

While Green *et al.* (2000) do not identify any studies that refute the feasibility of the TTO method, they do present some difficulties that have been encountered. In the study by Fryback *et al.* (1993) 38 respondents did not complete the TTO section of the interview. Of these, 28 refused to answer, finding the questions 'silly', 'too hard to imagine', or in conflict with personal philosophical religious beliefs, for other unexplained reasons. Five interviews were terminated by the interviewer for apparent problems in comprehension of the questions by the interviewees. The TTO section was skipped or terminated by the interviewer for five respondents who might have or did become distressed by the questions. Gage *et al.* (1996) asked 83 atrial fibrillation patients to value stroke related health states using TTO and SG. Results from 13 of the 83 subjects were excluded for one of three reasons: five subjects did not complete the interview because of difficulty with the utility assessment or time constraints, seven subjects did not understand one or more questions, and one subject said that she would rather die than have any health state, even her usual health. Unfortunately the manuscript does not make it clear whether these difficulties were predominantly encountered when respondents were faced with the TTO or SG questions.

Badia *et al.* (1999) assess the feasibility, reliability and validity of VAS and TTO. They ask 294 members of the Spanish population to value EQ-5D health states. Feasiblity was assessed in terms of percentage of non-response and missing values, difficulty ratings (interviewer and interviewee), and administration time. The non-response rate was very low for both VAS and TTO (three non-responses for VAS and five for TTO). Respondents found TTO more difficult than VAS. Twenty six percent of respondents thought VAS was either 'difficult or 'very difficult', while the corresponding figure for TTO was 37%. Interviewer ratings of difficulty suggested that very few respondents had difficulty with either of the methods (9% on the VAS and 8% on the TTO). Administration time was significantly longer for the TTO (21.8 minutes) than for the VAS (11.3 minutes). Badia *et al.* (2001) demonstrate the feasibility (and validity) of both the TTO and VAS methods for the valuation of temporary health states. The valuation of temporary health states is not an issue addressed in this thesis.

Wee *et al.* (2008) asked 62 members of the Singaporean general population to value EQ-5D health states using both SG and TTO. They assess the feasibility of the methods through a number of follow up questions. Respondents evaluated the following aspects of feasibility on a 10cm 0 to 10 VAS: 1) ease with which they understood the instructions; 2) ease with which they completed the exercise; and 3) amount of concentration needed. Subjects felt that both SG and TTO were easy to understand, with both methods receiving a mean score of 8 on the VAS (where 10 corresponds to easy and 0 to difficult). In response to the follow up question asking whether the TTO instructions should be amended six respondents felt the TTO instructions should be amended, while none thought the SG questions need to change. Of these six, four said they became confused and two felt that the term 'immediate death' was 'shocking' and that giving up life given by God was offensive. Respondents felt that both SG and TTO questions were easy to complete (both received a mean score of 8) and that they did not require a lot of concentration (both received mean scores of 3, where 0 corresponded to little concentration). When asked whether they preferred SG or TTO, 50% of respondents said they preferred SG, while 45% preferred TTO and 5% had no preference.

---Non-Trading

Both Fryback *et al.* (1993) and Handler *et al.* (1997) ask respondents to value their own health through TTO and encounter a high prevalence of non-trading behaviour i.e.

respondents are not prepared to give up anytime to improve their current health to perfect health. Fryback *et al.* (1993) find that the median TTO value is 1.00, meaning that more than 50% of respondents were not prepared to trade any time. This does not necessarily mean the TTO method was unfeasible. It could be that this many respondents were genuinely in full health. Handler *et al.* (1997) ask 100 respondents recruited from a hospital cafeteria (visitors and staff) to value their own health through TTO. While the interview failure rate was less than 3%, 24 of the 100 respondents were not prepared to trade any time. The authors found that those respondents unwilling to trade had less perceived control over their lives (as measured through Locus of Control Scale, Levenson 1974), although this was not statistically significant.

In the large scale study to derive the UK EQ-5D value set (Dolan *et al.* 1996b) 3395 respondents were asked to value 13 health states of differing severity. Although almost half of the respondents were prepared to sacrifice life expectancy in order to avoid all of the dysfunctional states they were asked to consider, one quarter were unwilling to sacrifice even a couple of weeks at the end of 10 years for 3 or more states. The explanation offered is a 'status quo' effect: people may give some epical status to their current position, and react asymmetrically to movements away from that position, placing greater weight on what they perceive as losses *vis-à-vis* the status quo in the form of reduced survival than on what they perceive as gains in terms of health status (in line with the theory of Kahneman and Tversky, 1979, discussed earlier). Therefore, it is suggested that there is some 'threshold of tolerability' that must be exceeded before respondents will give up time. This explanation is re-stated following qualitative follow-up with 43 of the respondents, as presented in Robinson *et al.* (1997).

Van Nooten *et al.* (2009) study determinants of willingness to trade in TTO exercises. They ask 339 respondents to value three EQ-5D health states through an internet questionnaire. For state 21211 78.2% of respondents did not want to give up any life years i.e. were non-traders. For state 22221 71.1% of respondents were non-traders and for state 33312 23% of respondents were non-traders. The study also collected age and subjective life expectancy. Probit regression analysis is performed in which the dependent variable is the willingness to trade, and the explanatory variables are remaining life expectancy (age minus subjective life expectancy), age, gender, education and own health. Remaining subjective life expectancy was the only significant variable, the effect being that the larger the remaining years of life, the lower the willingness to trade. The authors argue that this might be explained by the fact that respondents whose life expectancy exceeded 10 years felt 'cheated' out of life years. This might strengthen the affect of loss aversion mentioned earlier. Conversely if respondents are given more years than they had expected they would be more willing to give up years in a TTO.

In summary, there is strong evidence to support the feasibility and acceptability of the TTO method. Response and completion rates are high (typically in excess of 90%), and respondents commonly find TTO questions easier to complete than SG questions. One issue is the prevalence of non-trading. Evidence of non-trading among members of the general population valuing their own health is not concerning, but non-trading in valuations of poor hypothetical states is problematic. This may occur when respondents feel the 10 year time horizon is not realistic.

3.3.2.2 Reliability

(a) Internal Reliability (also known as Intra-rater reliability)

Torrance (1986) reviews the evidence on both the reliability and validity of the VAS, SG and TTO. Torrance considers internal reliability, test-retest reliability and precision. Internal reliability refers to a second measurement taken as part of the original interview. In Torrance's earlier work he found the coefficient of internal reliability to range from: 0.86 to 0.94 for the VAS (Torrance *et al.* 1982); 0.77 to 0.92 for the SG (Torrance 1976); 0.77 to 0.88 for TTO (Torrance, 1976; Torrance *et al.* 1982). No studies were identified that have used this measure of reliability to assess the TTO method since this early work by Torrance.

(b) Inter-rater reliability

Inter-rater reliability refers to the consistency in valuations between respondents. Froberg and Kane (1989) consider inter-rater reliability in their review, but identify only one study testing this (Patrick *et al.* 1973), and this study does not consider TTO. Krabbe *et al.* (1997) asked 104 students to value 13 EQ-5D health states, through SG, TTO, VAS, CV (they use the terminology 'WTP'), and the Paired Comparisons Method. The authors estimate inter-rater reliability though G-Theory. This method attributes variance to different sources and is based on a comparison with expected mean squares. Through this method the authors are able to attribute a proportion of the variance between individuals that is attributable to the health state. Coefficients are produced which the authors argue can be regarded as standard reliability coefficients. The coefficients were as follows: VAS (0.77), CV (0.49), TTO (0.65), SG (0.58). For the CV method more than 30% of the total variance was due to systematic differences between participants. Krabbe *et al.* (1997) argue that 'although a linear transformation of the mean WTP values to SG/TTO was technically possible, WTP in our operationalisation was found to be an inferior method with an unacceptably low reliability' (p.1649). They find the inconsistency between the dominant pairs of health states worrying and feel that WTP may have to be regarded as an unfeasible method. In the case of TTO and SG they conclude that the reliability was satisfactory.

(c) Test-retest reliability

Green *et al.* (2000) summarise the results of 12 empirical studies examining the test-retest reliability of TTO (Torrance, 1976; O'Connor *et al.* 1987; Reed *et al.* 1993; Ashby *et al.* 1994; Gabriel *et al.* 1994; Dolan *et al.* 1996a; Dolan *et al.* 1996b; Gage *et al.* 1996; Molzahn *et al.* 1996). The time between the initial test and the re-test in these studies ranges from 1 week to 1 year. The correlation coefficients range from 0.5 (Ashby *et al.* 1994) to 0.92 (Gage *et al.* 1996). Typically, the test re-test correlation coefficient is lower (suggesting poorer performance) the longer the time gap between the two valuations. For example, the study in which the time gap is 1 week (O.Connor *et al.* 1987) produces a correlation coefficient of 0.87, while the study in which the time gap is 1 year produces a coefficient of 0.62 (Torrance, 1976). Torrance (1986) suggests this may, in part at least, be explained shifts in people's preferences over time. Alternative explanation is that when the time gap is short respondents are able to remember the valuation that they gave in the first instance. There are five studies which directly compare SG and TTO, and one study uses both props and no props versions giving a total of 6 points of comparison (Torrance 1976; O'Connor *et al.* 1987; Reed *et al.* 1993; Dolan *et al.* 1996b; Gage *et al.* 1996). TTO outperforms SG in three of five points of comparison (in one case a range of coefficients is given for TTO making it difficult to determine whether it outperformed SG). Ultimately, Green *et al.* (2000) argue that VAS, SG and TTO all 'demonstrate an acceptable level of reliability' (p.157).

Further evidence on the test-retest reliability of the TTO method has been published since the review by Green *et al.* (2000). Badia *et al.* (1999) perform test-retest analysis on 50 respondents. These 50 respondents repeated the interview one to four weeks after the initial administration, with the same interviewer performing both interviews, and the states being valued in the same order. While the studies identified by Green *et al.* (2000) typically use Pearson correlation coefficients, Badia *et al.* also estimate reliability coefficients using Generalizability Theory. Badia *et al.* (1999) estimates three main sources of variance: 'health state', 'individual' and 'time'. The most reliable method is that with the lowest proportion of variance attributable to time. The ICC was 0.90 for VAS and 0.84 for TTO. This is similar to the results reviewed in Green *et al.* (2000) in the studies in which the time gap was similar. There were only significant differences between the first and second TTO valuations for one state (21232), while there were significant differences in five states for VAS. The results of the generalizability analysis showed that the variability attributable to time was zero for both methods. Badia *et al.* (1999) conclude that both methods demonstrated high test-retest reliability at individual and aggregate level.

Groome *et al.* (1999) ask 64 Canadian patients with renal failure to value a number of renal related treatments through VAS, TTO and SG. Ten days later they were asked to repeat the valuations. The reliability between the valuations is assessed through the coefficient of repeatability: the maximal difference one would observe between assessments from once occasion to the next within an individual and is calculated by multiplying the standard deviation of the differences by 1.96 (Bland and Altman, 1986). The coefficient of repeatability was 27.4 for SG, 38.4 for TTO and 36.5 for VAS. This indicates that the SG values could differ by as much as -27 to +28 (on a scale from 0 to 100) from one occasion to the next, while the TTO values could differ by as much as -42 to +35. The authors argue that 'the results of this study cast doubt on the repeatability of the SG, TTO and VAS methods, which is a necessary element in the determination of the validity of a measurement method' (p.856). However, in the context of economic evaluation what matters is the aggregate values. At the aggregate level there were no statistically significant differences between the first and second valuations for any of the methods.

(d) Other assessments of reliability

Torrance (1986) considers reliability in terms of the precision of an individual measurement, where precision is determined by the standard deviation (SD) of the measurement error (σ). Torrance (1976) found the SD to be 0.13 for both SG and TTO, while Torrance *et al.* (1982) found the SD to range from 0.09 to 0.15 for VAS. Therefore, if an individual responds to a TTO question with a utility of 0.60 for a particular health state, the 95% confidence interval would be 0.34- 0.86. Arguably, this assessment of reliability is of limited use, since the imprecision of individual measurements can be ameliorated by taking the mean of a large group of subjects. Since the standard error of the mean is $\sigma/\sqrt{N}$, the mean utility value for a health state can be made as precise as desired by increasing the group size N. Therefore, the above results suggest that a larger sample size will be required to produce accurate results for VAS than will be required for SG or TTO.

In summary, the evidence suggests the TTO method is reliable. Only the study by Groome *et al.* (1999) questioned the reliability of the method, but this study found no significant differences between the first and second valuations at the aggregate level.

3.3.2.3 Validity

Green *et al.* (2000) assess the validity of SG, TTO and VAS in a number of ways. They first consider the theoretical basis of the methods, before considering any empirical support for these theoretical underpinnings. They then consider how accurately the methods predict preferences.

3.3.2.3.i Theoretical Basis (Internal Validity)

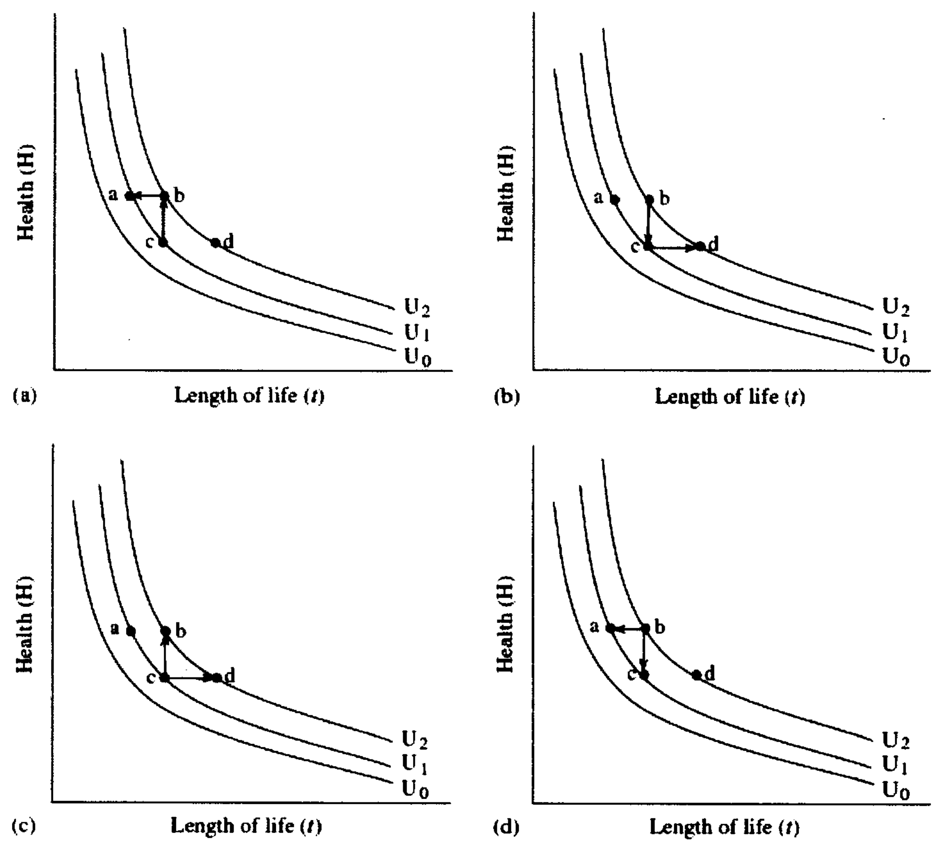
The SG method is strongly rooted in von Neumann-Morgenstern utility theory (von Neumann and Morgenstern, 1944) and as such has been referred to as the 'gold standard' of preference elicitation (Torrance, 1976; Torrance and Feeney, 1989). However, it should be noted that this position has been challenged (see for example Loomes and Sugden, 1982; Kahneman and Tversky, 1989; Torrance and Feeney, 1989; Richardson, 1994). In contrast to

SG, the TTO method arose not from any particular theory of utility, but rather from a pragmatic desire to generate valuations with similar empirical properties to SG, while offering a more feasible alternative to it (Buckingham and Devlin, 2006). Therefore, since TTO valuations do not conform to the utility-under-uncertainty requirements of von Neumann and Morgenstern, they have been considered by some to be "values" rather than utilities *per se* (Drummond *et al.* 2005), though this interpretation has been contested (Richardson, 1994).

Attempts have been made in the literature to place TTO within a theoretical framework. Mehrez and Gafni (1990) show TTO choices using indifference curves and discuss TTO in the context of value function theory. Value function theory was initially developed (Dyer and Sarin, 1982) to examine strength of preference when making decisions under uncertainty. Mehrez and Gafni (1990) apply this theory to the choice between quality and length of life in TTO exercises. Essentially they argue that TTO valuations can not be compared unless the underlying value function has been validated. This is not an area that has been pursued further in the literature, and Mehrez and Gafni make no recommendation as to how the value function could be validated.

Buckingham *et al.* (1996) were the first to portray TTO in the context of welfarist theory and Hicksian compensating variation. In this context TTO appears not too dissimilar to CV. To reiterate, compensating variation measures a welfare gain by the compensating loss of something else that is valuable, so that the respondent is returned to his/her original level of welfare. In the context of TTO the welfare gain results from an improvement in health, and the compensating loss is in the form of reduced life expectancy. The respondent must decide what reduction in life expectancy will return them to their original level of utility. This potential theoretical framework is developed further by Buckingham and Devlin (2006). They posit four distinctive TTO approaches, one of which is the standard TTO method outlined above. The standard TTO approach elicits a respondents compensating variation for a health gain. However, in the context of Hicksian theory (Hicks, 1943; as outlined in the introduction in the context of contingent valuation) there are three further valuation possibilities: compensating variation of a loss, equivalent variation of a gain, equivalent variation of a loss. The four different valuation possibilities are illustrated in figure 3.1.

Figure 3.1: (a) compensating variation for a gain; (b) compensating variation for a loss; (c) equivalent variation for a gain; (d) equivalent variation for a loss (taken from Buckingham and Devlin, 2006, p.1151).



In each of the graphs in figure 3.1 the x axis represents length of life (t), and the y axis represents health (H). U_0, U_1 and U_2 represent indifference curves that portray a given level of utility that can be achieved from different combinations of length of life and health, where $U_2>U_1>U_0$. Figure 3.1 (a) represents the conventional TTO approach (compensating variation for a health gain). Movement from c to b represents the health gain, while movement from b to a represents the amount of time that must be lost to return the individual to his/her original level of utility (U_1). Figure 3.1 (b) represents the compensating variation for a health loss. Movement from b to c represents the fall in health, while the movement from c to d represents the increase in life expectancy that is required to return the individual to their original level of utility (U_2). Figure 3.1 (c) represents the equivalent variation for a gain. The movement from c to b represents the health gain, and the movement from c to d represents the increase in life expectancy that would take the respondent to the same level of (higher) utility (U_2) generated by the increase in health.

Figure 3.1(d) represents the equivalent variation for a loss. The movement from b to c represents the fall in health, while the movement from b to a represents the fall in life expectancy that would take the respondent to the same level of (lower) utility (U_1) generated by the fall in health.

Considering TTO valuations as Hicks-like measures of welfare change throws the focus on the relation between values and the marginal rate of substitution between quality and length of life, and also allows widely discussed biases (addressed later) to be interpreted in terms of the shape of the underlying indifference curve. Buckingham and Devlin (2006) identify three formulations of TTO that have been tested in the literature, all of which can be considered to elicit the compensating variation for a loss. Spencer (2003) asked subjects to imagine living in a health state for a period of two years (this state having already been valued through the conventional TTO), followed by death. They were then asked to trade by prolonging life in a lower quality health state. This corresponds to figure 3.1(b) above. The value obtained through the unconventional method was calculated as follows. Say, for example, through the conventional method 10 years in state 21222 was considered equivalent to 6 years in full health. Then in the unconventional approach 2 years in state 21222 was considered equivalent to 3 years in state 22232, the value for 22232 would be given by:

$$2 * 0.6 = 3 * V(22232)$$
$$V(22232) = 1.2/3 = 0.4$$

Three EQ-5D health states were valued through both the conventional and unconventional TTO methods: 21222, 22232, 21211. The value obtained for state 21222 was considerably lower than the value obtained through the conventional compensating variation for a gain approach.

This framework is revisited in the following chapter as it is central to the theoretical basis of the exploratory method that is developed.

3.3.2.3.ii Discriminant Validity

Spencer (2003) identified four issues that influence a respondent's preferences that are not typically taken in to account when estimating health state values from TTO responses: a) maximal endurable time (MET), b) time preference c) strong preference for longevity of life (scale compatibility), d) loss aversion. The first two points are possible explanations for a violation of Constant Proportional Time Trade-Off (CPTTO), while the other two are separate sources of bias. Another potential source of bias when health profiles are being valued is a failure of the assumption of additive separability.

(a) Constant Proportional Time Trade-Off (CPTTO)

Tsuchiya and Dolan (2005) identify 10 studies that test the assumption of CPTTO, (Sackett and Torrance, 1978; Pliskin *et al.* 1980; McNeil *et al.* 1981; Miyamoto and Eraker, 1988; Cook *et al.* 1994; Buckingham *et al.* 1996; Dolan *et al.* 1996b; Bleichrodt and Johannesson, 1997; Stalmeier *et al.* 1997; Unic *et al.* 1998). Many of the studies have shown CPTTO to hold at the aggregate level, while violations occur at the individual level. Pliskin *et al.* (1980) asked 10 respondents the number of years they would sacrifice to avoid severe or mild angina using both 5 and 15 year durations. At the individual level most respondents violated CPTTO, while at the aggregate level there was little difference between the trade-offs in the 5 and 15 year durations. Cook *et al.* (1994) used gallstone disease states lasting 12 months and 12 years and found no significant differences at the aggregate level. Bleichrodt and Johannesson (1997) obtained the same result using durations of 10 years and 30 years. Other studies have shown CPTTO to be violated at the aggregate level. Typically they find that respondents trade off proportionally less time when the duration is shorter (hence giving a higher health state value). Sackett and Torrance (1978) asked dialysis patients and members of the public to value various health conditions with durations of 3 months, 8 years and the life expectancy of the respondent. They found that values declined with duration. Miyamoto and Eraker (1988) found that respondents did not trade off any time to improve their current health when the duration was under 1 year, whereas time was traded off when the duration was more than 1 year.

Buckingham *et al.* (1996) test three forms of TTO for a condition that lasts the rest of one's life. These were a daily TTO, which was about trading off the number of hours awake per day; a yearly TTO, which asked respondents to trade off the number of active days per year; and a lifetime TTO, which asked respondents to trade off years of life expectancy. They found that the yearly values were the highest and the daily values were the lowest, suggesting the relationship between duration and health state valuations is not constant.

Dolan (1996a) has found VAS valuations to be susceptible to duration. Tsuchiya and Dolan (2005) identify the articles that study the affect of duration on SG valuations (McNeil *et al.* 1981; Bleichrodt and Johannesson, 1997; Bala *et al.* 1999). All three studies show that utility independence is violated at the aggregate level. This shows that the affect of duration on health state valuations is not exclusive to the TTO method.

Attema and Brouwer (2010) update the review by Tsuchiya and Dolan (2005) to include studies published since 2002, and to include correction for utility curvature. They identify eight articles not included by Tsuchiya and Dolan (Hall *et al.* 1992; Stiggelbout *et al.* 1995; Kirsch and McGuire, 2000; Martin *et al.* 2000; Stalmeier *et al.* 2001; Bleichrodt *et al.* 2003; Dolan and Stalmeier, 2003; van der Pol and Roux, 2005). Dolan and Stalmeier (2003) is discussed below under MET, Bleichrodt *et al.* (2003) is considered under Loss aversion, van der Pol and Roux (2005) and Martin *et al.* (2000) are considered under time preference. Hall *et al.* (1992) compared three different durations, 10, 50 and 100% of life expectancy. No violations of CPTTO were found at the aggregate level. Stiggelbout *et al.* (1995) used short and intermediate durations and found a violation of CPTTO with TTO values for short durations being higher than those for long durations (aggregate). Kirsch and McGuire (2000) compared a short and an intermediate duration and found mixed evidence. Stalmeier *et al.* (2001) found smaller TTO values for longer durations when comparing two intermediate durations in a severe health state. Attema and Brouwer conclude that 'it appears difficult, therefore, to derive any definite answers from the literature regarding CPTTO' (p.493). Attema and Brouwer (2010) perform their own experiment and find that CPTTO is violated at the aggregate level.

Table 3.1 summarises the evidence on CPTTO. Note this table is similar to one presented in Attema and Brouwer (2010), but the table presented here includes one additional study (Buckingham *et al.* 1996) which seems to have been overlooked in the review by Attema and Brouwer. Of the 18 studies identified, ten have found evidence of violation of CPTTO at the aggregate level. Of these ten studies, seven have found that shorter durations elicit higher valuations.

(b) Explanations for violation of CPTTO

---Maximal Endurable Time (MET)

If the health state is particularly severe a respondent may doubt their ability to cope with continuing ill-health. Tsuchiya and Dolan (2005) consider the QALY model and its assumptions. One of the assumptions they consider is whether or not the value of a health state is affected by how long the state lasts. Within this, they consider utility independence, CPTTO and MET. The only articles they identify addressing MET are the ones by Sutherland *et al.* (1982) and Stalmeier *et al.* (1996). Sutherland *et al.* (1982) asked 20 physicians their preferences for 5 health scenarios. Preferences were elicited through a 'preference questionnaire', which asked respondents whether they preferred a defined time in the specified scenario to immediate death. The second method, the authors describe as a 'certainty equivalence approach', and is essentially an SG procedure. The results of the preference questionnaire suggested there were changes in attitude toward survival in some of the health states as the amount of time to be spent in the state increased and/or the state became more dysfunctional. For the worst health scenario, as the time to be spent in the scenario increased from 3 months to a lifetime, the proportion of raters who stated a preference for the scenario over immediate death decreased from 60% to 10%. The results of the 'certainty equivalence' exercise revealed a preference for shorter duration of survival in the more dysfunctional health states. Ultimately, the evidence suggests that when the time spent in a dysfunctional health state exceeds a particular duration (which varies by individual) any additional time spent in that state was regarded as a penalty, and assigned a negative value relative to death. This result suggests the results of both SG and TTO preference elicitations for severe health states may be influenced by the duration respondents are expected to endure the hypothetical scenarios for.

Table 3.1 Overview of Studies Considering Constant Proportional TTO

Study	**Durations Used**	**CPTTO violated at the aggregate level?**	**Which durations gave higher values?**
Sackett and Torrance (1978)	3 months, 8 years and life expectancy	Yes	Shorter durations
Pliskin *et al.* (1980)	5 and 15 years	No	
Miyamoto and Eraker (1988)	<1 year and >1 year	Yes	Shorter durations
Hall *et al.* (1992)	10%, 50% and 100% of life expectancy	No	
Cook *et al.* (1994)	12 months and 12 years	No	
Stiggelbout *et al.* (1995)	3, 10 and 15 years	Yes	Shorter durations
Buckingham *et al.* (1996)	1 day, 1 year, lifetime	Yes	1 year
Bleichrodt and Johannesson (1997)	10 and 30 years	No	
Stalmeier *et al.* (1996)	5, 10, 25, 50 years	No	
Stalmeier et al. (1997)	5, 10, 25, 50 years	Yes	Shorter durations
Unic *et al.* (1998)	5 and 10 years	Yes	Longer durations
Kirsch and McGuire (2000)	2 and 10 years	Yes	Shorted durations
Martin *et al.* (2000)	5, 10 15 years	Yes	Shorter Durations
Stalmeier *et al.* (2001)	10 and 20 years	Yes	Shorter durations
Blecihrodt *et al.* (2003)	13, 19, 24, 31, 38 years	No	
Dolan and Stalmeier (2003)	10 and 20 years	No	
van der Pol and Roux (2005)	20 and 50 years	No	
Attema and Brouwer (2010)	14 and 27 years	Yes	Longer durations

Stalmeier *et al.* ask 86 respondents whether they would rather live for 25 years or 50 years in a breast cancer state. Fifty-eight percent indicated MET such that 25 years with metastasized breast cancer was preferred to 50 years in the same state (even though the state was valued as better than dead). However, the TTO valuations were severely inconsistent with the preference for the shorter life duration and instead complied with a CPTTO (i.e. approximately twice as many years are traded in the 50 year duration as in the 25 year duration). Stalmeier *et al.* consider these inconsistencies to be 'preference reversals', and believe them to be driven by the blind application of a 'proportional heuristic' (respondents simply apply the same proportions regardless of the duration of the TTO exercise, even though they may in fact prefer one duration to another).

Stalmeier *et al.* (2007) asked 123 Dutch respondents to state whether they preferred given EQ-5D health states of differing durations to death. The respondents were also asked to choose between short and long durations in these health states. The results showed that for moderate and severe health states the proportion of respondents who considered a state to be better than dead decreased as the duration of the health state increased, suggesting an interaction between health and duration. Dolan and Stalmeier (2003) use a 10-year and 20-year TTO to value an EQ-5D health state. They find that CPTTO holds, both for respondents who report MET preferences, and for those who do not. For respondents who report MET preferences, the results suggest a proportional heuristic is being applied. The authors draw the following conclusion, 'for more severe states, such as those containing one or more dimensions at the third and most severe level on the EQ-5D, it might be misleading to use the tariff to calculate QALYs in that such states might be associated with MET preferences. Therefore, a general set of health state values should be used with caution in a cost utility analysis where the health effects are expressed in terms of QALYs" (p.455).

---Time Preference

Like MET, time preference is a potential explanation for the violation of CPTTO. Respondents may give greater weight to benefits (e.g. good health) occurring tomorrow than in ten years time, for example. This would not only represent a violation of the assumptions of the QALY model, but would also lead to a source of double counting, since the benefits in cost effectiveness analysis are typically discounted (NICE recommends a rate of 3.5%, NICE 2008). If an individual has a positive rate of time preference, he/she will be prepared to sacrifice more years of life in the future relative to years of life now, which will lead to a downward bias in utility values (see Dolan and Jones-Lee, 1997).

Van der Pol and Roux (2005) identify a number of studies that test for the existence and extent of time preference in TTO exercises. The most common method of doing this is by estimating the time preference effect from differences in TTO values for different durations (Olsen, 1994; Gyrd-Hansen, 2002; Stavem *et al.* 2002). These studies have produced mean time preference rates ranging from 0.07 to 0.10. Dolan and Gudex (1995) use a slightly different approach. They ask 39 members of the general population to value five EQ-5D health states for duration of one month, one year and ten years, using TTO. The 10 year TTO was followed by immediate death, while the shorter durations were supplemented with time in full health up to a total of 10 years, which would then be followed by death (the authors felt shorter life expectancies would be unreasonable). The procedure was also varied so that the one year in poor health could occur either at the start or the end of the 10 year duration (the one month duration always occurred at the start). Considering the results from the two different one year duration protocols, only one quarter implied a positive discount rate, while 39% implied a negative discount rate (the remainder implied a zero discount rate). In addition there was inconsistency within respondents e.g. one third of respondents displayed a positive rate of time preference for some states a negative rate for other states. Ultimately, at the aggregate level, the results suggest an indifference towards the timing of poor health. The authors conclude that the 'implicit assumption of the TTO method, that the rate of time preference is zero, is valid at the aggregate level' (p.296).

Another approach involves testing the convergent validity of TTO values adjusted for time preference with other measures of health state preference (Stiggelbout *et al.* 1995; Martin *et al.* 2000). One method of adjustment involves identifying an individual's utility function for life years by eliciting certainty equivalents for gambles with different life expectancies (Sox *et al.* 1986). The second method adjusts TTO values for time preferences by dividing discounted life years in full health by discounted years in the health state. The size of the adjustment is a function of the time preference rate, the duration of the health state, and the size of the unadjusted TTO value (Johanneson *et al.* 1994). Van der Pol and Roux (2005) apply a version of the latter adjustment procedure. They ask respondents to value a weight gain health scenario using both a 20 year TTO and a 50 year TTO. They also ask a separate time preference question. The question asks respondents to specify the number of years with their weight being 20% higher starting in 45 years time that is just as bad as their weight being 20% higher for 5 years starting in 15 years. The majority of respondents expressed a positive time preference (63%). The sample completing the 20 year duration TTO questions had a mean time preference rate of 0.024, while the sample completing the 50 years duration questions had a mean time preference rate of 0.009. The mean TTO values were higher for the 20 year duration questions than the 50 year duration questions, although there were no significant differences between the two for either the unadjusted values or the values adjusted for individual time preference. The mean unadjusted TTO values were 0.677 for the 20 year duration and 0.658 for the 50 year duration. The values adjusted for individual preferences were 0.717 for the 20 year duration and 0.686 for the 50 year duration.

In summary, it seems that CPTTO may not hold. The two foremost explanations for this are MET and time preference. Evidence on both is mixed. MET would only seem to be relevant for particularly poor health states. The identified rates of time preference ranged from 0.009 to 0.10. Ways of adjusting for time preference have been explored but these are yet to be strongly validated.

(c) Other Sources of Bias

---Loss Aversion

Bleichrodt (2002) and Spencer (2003; 2004) consider the impact on TTO valuations if a respondent evaluates the question as gains and losses relative to a reference point. Central to this notion is the assumption that a respondent is more sensitive to losses than to gains, termed loss aversion (Kahnemann and Tversky, 1992). The second assumption is that there is diminishing sensitivity to gains or losses as these increase from the reference point. Hence, a respondent gives relatively greater emphasis to small-to-medium gains and losses than they do to larger gains and losses. Both Bleichrodt and Spencer assume that the reference point is the initial health state that is considered in each question. Bleichrodt

(2002) explains how loss aversion leads to upward bias in the TTO valuation. A gain in health status is offset by a loss of duration. In the presence of loss aversion the respondent will give up fewer years to reach indifference i.e. a smaller loss is required to produce the necessary utility decrement to offset the utility increment generated by the gain in health status.

Bleichrodt *et al.* (2003) perform an empirical test of loss aversion (Note, this is also a test of procedural invariance i.e. a test of whether the values given are influenced by the particular procedure employed). They first ask 51 students to complete a conventional TTO exercise valuing back pain for durations of 13, 19, 24, 31 and 38 years. They then ask them to complete an alternative TTO exercise in which the time in full health is held constant (at the level obtained in the conventional TTO) and the time in the impaired health state is increased to reach indifference (compensating variation for a health loss seen in figure 3.1b earlier). For example, using duration of 19 years, if the respondent in the conventional exercise felt that 10 years in full health was equivalent to 19 years with back pain, in the unconventional exercise the respondent would be asked how many years with back pain they thought was equivalent to 10 years in full health. In the absence of loss aversion the respondent should reach indifference at 19 years, so that the same utility value is elicited through both procedures. The results showed that for the shorter durations (13, 19 and 24 years) the conventional TTO utilities were statistically significantly higher than the unconventional TTO utilities. When comparing the conventional TTO values by duration the evidence is mixed. However, when the unconventional TTO values are compared by duration CPTTO is violated.

Attema and Brouwer (2008) also performed a test of procedural invariance using the conventional and unconventional TTO questions above, and found further evidence that the conventional TTO questions elicit higher values than the unconventional questions. However, further testing by Attema and Brouwer (2011) found little evidence of any violation of procedural invariance.

---Scale Compatibility

Scale compatibility means that an individual assigns more weight to an attribute the higher its compatibility with the response scale used (Bleichrodt, 2002). A theory of scale compatibility is described by Tversky *et al.*(1988), and Delquie (1993; 1997) presents empirical evidence of the impact of scale compatibility on individual preferences. In TTO the response scale is duration i.e. the respondent trades in time, not health status. Scale compatibility then implies that the individual will give more weight to duration than to health status. This means that respondents influenced by scale compatibility will be prepared to trade fewer years to achieve full health than if they were not influenced by scale compatibility. Therefore scale compatibility causes an upward bias in TTO valuations. Extensive empirical evidence is presented in Delquie (1993; 1997) and Bleichrodt and Pinto find scale compatibility in medical trade-offs (2003).

---Additive Separability

An assumption of the QALY model is that the value of a health state should be independent of what precedes or follows it (Broome, 1993). This is the assumption of 'additive separability' and it means that the value of a complete health profile would be equal to the sum of the value of the individual health states that make up that profile, irrespective of the order of the states. Tsuchiya and Dolan (2005) consider the evidence on whether or nor this assumption holds. They identify five studies testing this assumption, three of which use the TTO method (Richardson *et al.* 1996; Krabbe and Bonsel, 1998; MacKeigan *et al.* 1999). Richardson *et al.* (1996) ask women to value breast cancer related health scenarios using VAS, SG and TTO. Three scenarios consisted of a single health state, whereas the last one was a profile combining these 3 states in deteriorating order followed by death. The found that the number of QALYs calculated indirectly from the individual health states was 30% to 50% higher than the number of QALYs calculated from the direct value of the profile.

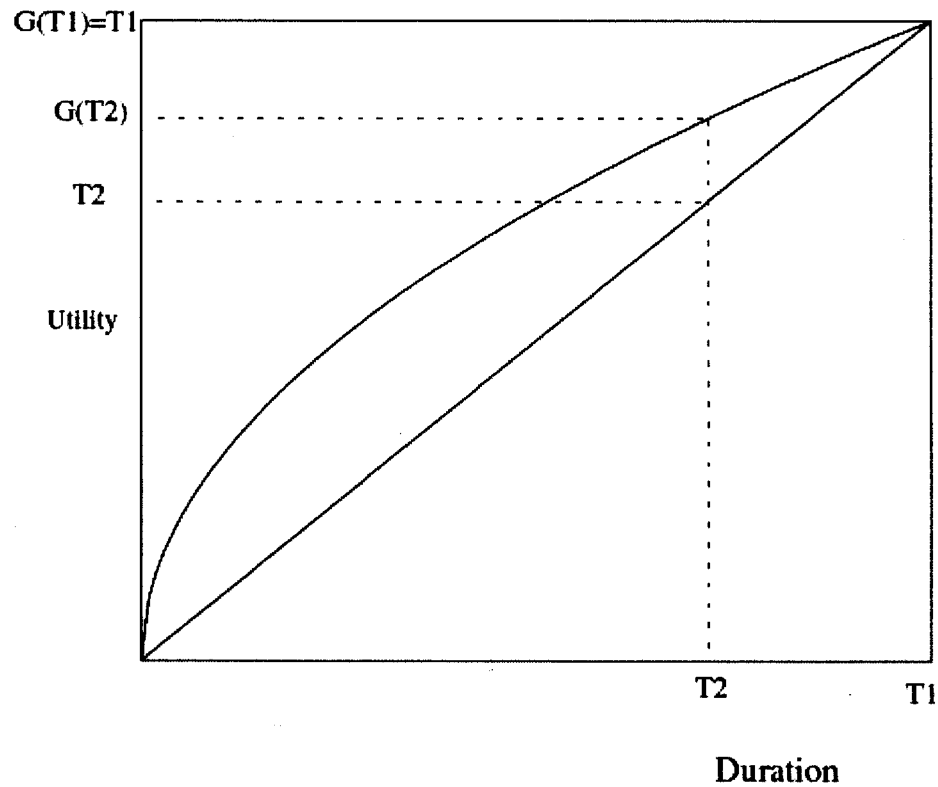
Krabbe and Bonsel (1998) asked 104 students to value EQ-5D states using TTO on two occasions. On the first occasion respondents are asked to choose between living in a fixed EQ-5D state for 10 years, or living for x years in the 'best imaginable health state' followed by (10-x) years in the 'worst imaginable health state'. On the second occasion the choice is between 10 years in the EQ-5D health state, and z years in the 'worst imaginable health state' followed by (10-z) years in the best imaginable health state (i.e. the ordering is reversed). In order for the assumption of additive separability to hold the number of years spent in the best health state should be equal in the two scenarios (given appropriate discounting). When a discount rate of 5% was applied the assumption of additive separability held for two thirds of respondents. However, it is difficult to know if this discount rate accurately reflects respondents time preference. If it does not then two thirds may be an overestimate of the proportion of respondents meeting additive separability. Of the remaining one third, the majority preferred to live in the worst health state first.

MacKeigan *et al.* (1999) asked type 2 diabetic patients to value various treatments through VAS and TTO. Combination therapies were valued both directly and indirectly and no significant differences were found between the two, meaning the assumption of additive separability was not violated. More recently, Bleichrodt and Filko (2008) used SG type questions to test Additive Separability using Generalized Marginality, which they believe avoids the need for some of the assumptions made by the studies above such as constant discounting. Through this test they argue that while Additive Separability is violated at the individual level, it cannot be rejected at the aggregate level. The approach has been criticised (Gandjour and Gafni, 2010), and the originators have subsequently sought to defend their method (Bleichrodt and Filko, 2010).

3.3.2.3.iii Convergent Validity

If one believes in the strong theoretical basis of the SG method, a validity test of the TTO method would be the similarity of the results produced by the two methods. A number of empirical studies have shown the TTO and SG methods to yield systematically different results (Torrance, 1976; Read *et al.* 1984; Hornberger *et al.* 1994; Stiggelbout *et al.* 1994; Bleichrodt and Johannesson, 1997; Lenert *et al.* 1998). The common pattern is that SG utilities exceed TTO utilities. The traditional explanation for this pattern is that respondents have a concave utility function for duration (McNeill *et al.* 1978; Stiggelbout *et al.* 1994; Chapman, 1996; Verhoef *et al.* 1994; Wakker and Deneffe, 1996; Cairns and van der Pol, 1997; Bleichrodt and Pinto, 2000). By assuming a linear utility function the TTO method is assuming risk neutrality. If respondents are risk averse the appropriate utility function is concave. To understand why utility curvature causes a downward bias in TTO utilities consider the graphical analysis presented in figure 3.2, as outlined by Bleichrodt (2002). The x axis represents duration in a given health state (Q1), and the y axis represents the utility derived from being in that state for a given duration. The utility function over duration (G) is scaled such that $G(T_1)=T_1$. $G(T_2)$ represents the utility derived from duration T_2 when the utility function is concave (compared to T2 when the function is linear). Since $G(T_2)>T_2$ by the concavity of G, it follows that $G(T_2)/G(T_1)>T_2/T_1$. Thus, the true utility of health state Q_1, $G(T_2)/G(T_1)$, exceeds the TTO utility, T_2/T_1, or the TTO utility is biased downwards. Since the SG method imposes no restrictions on G (see Bleichrodt, 2002), the utility curvature does not lead to bias in the SG utilities.

<u>Figure 3.2 – Concavity of the utility function implies that the TTO utilities are biased upwards (Bleichrodt, 2002)</u>



Authors, such as Miyamoto and Eraker (1985), have suggested adjusting TTO values to account for risk preferences. Stiggelbout *et al.*(1994) ask testicular cancer patients to value two hypothetical testicular cancer related health states (each of two durations, two and 10 years) through TTO, SG and Certainty Equivalent (CE) questions. CEs ask respondents the number of years in good health for certain that he or she considers equivalent to a gamble involving an X% chance of a long and a (100-X)% chance of a short length of life in good health as best and worst outcomes respectively (Keeney and Raiffa, 1976). Subjects who value nearby years higher than years further away will accept CEs that are less than the expected value of the gamble (i.e. they display risk aversion). Stiggelbout *et al.* (1994) adjust the TTO values they obtain using the risk preferences elicited from the CEs. The results show that the SG values were significantly higher than the unadjusted TTO scores for all profiles. However, adjusted scores were higher than unadjusted scores, and were not significantly different from those obtained from the SG for three of the four profiles.

Bleichrodt (2002) has argued that unadjusted TTO scores may be more consistent with individual preferences than adjusted TTO scores. Bleichrodt considers three sources of bias affecting TTO: utility curvature, loss aversion and scale compatibility, and three sources of bias affecting SG: loss aversion, scale compatibility and probability weighting. Loss aversion and scale compatibility in the context of TTO were outlined earlier, and both produce upward biases. Utility curvature, as outlined above, causes a downward bias in TTO utilities. Therefore, Bleichrodt argues the TTO method can both over- and underestimate utility depending on the relative size of the biases.

Probability weighting arises when respondents do not evaluate probabilities linearly in the SG method. Empirical evidence shows that the probability weighting function is typically inverse S-shaped (Lattimore *et al.* 1992; Tversky and Kahneman, 1992; Camerer and Ho, 1994; Bleichrodt and Pinto, 2000; Abdellaoui, 2000). This means that respondents overweight small probabilities and underweight larger probabilities. The point where the function changes from overweighting probabilities to underweighting them lies at approximately 0.35 (Bleichrodt, 2002). Underweighting leads to a utility value that is an overestimate. Since the probabilities that are reported in SG elicitations are generally well above 0.35 (e.g. Lenert *et al.* 2001) probability weighting will tend to cause an upward bias of utilities elicited through the SG method. Loss aversion causes an upward bias in SG utilities as it does in TTO utilities. The effect of scale compatibility on SG utilities is ambiguous depending on which of the three probabilities the respondent focuses on. If the respondent focuses on the probability of death there will be an upward bias in the SG utility; if the respondent focuses on the probability (p) of the good outcome there will be a downward bias; while focusing on the probability of the poor outcome (1-p) leads to an upward bias (see Bleichrodt, 2002, for further explanation). Therefore, given that two of the three sources of bias in SG utilities are upward, and the other is ambiguous, Bleichrodt argues SG utilities are generally biased upwards. Bleichrodt (2002) concludes that 'the assumption of linear utility, which is often believed to be a weakness of the TTO, is crucial in the explanation for why the TTO can be more consistent with individual preferences than the SG. Without this assumption the TTO would also be biased upwards due to loss aversion and scale compatibility. This observation implies that proposals to adjust TTO measurements for utility curvature may actually decrease the consistency of the TTO with individual preferences' (p.454).

Another way of assessing convergent validity is consistency with rank ordering. Ashby *et al.* (1994) ask various samples (nurses, hospital doctors, GPs and members of university staff) to value five health states following treatment for breast cancer. The respondents first ranked the states before valuing them through the TTO method. The authors found that the TTO valuations were consistent with the rank orderings of the states. Dolan *et al.* (1996a) ask respondents to value EQ-5D states using VAS, TTO and SG. Since some EQ-5D states are logically ordered with respect to others the authors were able to test the consistency of the methods by assessing whether the implied rankings met logical consistency. They found that the TTO method produced more consistent results than the SG method (results are not presented for VAS). The TTO method produced a consistency rate of 91.7% (n=145), compared with 83.8% (n=136) for SG. Gage *et al.* (1996) found that TTO values reflected the expected ordinal ranking of stroke severity (mild, moderate, severe) and the TTO and SG values for moderate stroke were not significantly different from each other (SG was only used to value moderate stroke). In other studies Dolan *et al.* (1996b) have reported high levels of consistency amongst a large general population sample, and Churchill *et al.* (1984, 1987) determined that patients' measured utilities (for own health) correlated significantly with nephrologists' ratings of the patients' quality of life.

Robinson *et al.* (1997) attempted to test the validity of the TTO method by eliciting qualitative data from a subset of the MVH sample. Respondents who rate the same state as better than dead on the VAS but worse than dead in the TTO were asked whether or not the position of that state on the VAS meant that they personally preferred spending 10 years in that state to immediate death. Of the 29 respondents that fell into this category 12 said that it did, 14 said it did not, whilst three did not know. It appeared that certain respondents were ignoring the duration of the state when completing their VAS. Thirteen of the 29 respondents with a different ordering of dead between the VAS and TTO said that the 10 year time scale had been more salient in the TTO than in the VAS. Of the 29 respondents, 18 had put dead at the bottom of the VAS scale. Of these 14 made comments which seemed to imply that for them death "automatically" went to the bottom of the scale. Respondents were asked whether they would still consider a state to be worse than dead in the TTO if it lasted for the rest of their life rather than for 10 years. Only three of the 29 respondents changed their preference, suggesting that any 'dread' concerning the time of death was not a significant factor.

Fifteen of the 43 respondents refused to trade off even a few days or weeks in order to avoid a health state which they had placed below 11111 on the VAS. All 15 confirmed that their VAS response did mean that they considered 10 years in that state to be worse than 10 years in state 11111. However, they did not translate this into a willingness to trade off time to avoid that state. Only one respondent refused to trade-off time throughout the TTO exercise suggesting this finding is not a result of respondents refusing to 'play the game'. The authors argue there is a 'threshold of tolerability' below which states have to fall before a respondent is willing to give up any time. This may be explained by the loss aversion concept mentioned earlier i.e. a disproportionately large gain in improved health status is required as compensation for the loss of life expectancy.

In summary, evidence suggests the TTO and SG methods do not produce equivalent results. Since the SG method originates from the principles of utility theory it could be argued that the TTO method must be biased. However, as argued by Bleichrodt (2002) the TTO method may actually be the more accurate of the two. Sources of bias in the TTO method (scale compatibility, loss aversion and utility curvature) may balance out to a greater extent than the sources of bias in the SG method (scale compatibility, loss aversion and probability weighting). Therefore, the differences in values between the two methods may be cause by an upward bias in SG responses.

Comparison of TTO and VAS shows that the TTO method is insensitive to very small benefits, which is consistent with the prevalence of non-trading seen in the section on feasibility.

3.3.2.4 Summary of findings on the TTO method

The feasibility of the method is good, with respondents reporting good ease of completion. Non-trading can be an issue if the health gain is small (such as from own health to perfect health in general population samples). The evidence supports the reliability of the method. Although the method did not arise from a particular theoretical basis (as is the case with SG) it has been demonstrated that responses can be interpreted in a Hicksian compensating variation framework. Finally, while the evidence seems to suggest that constant proportional time trade off does not hold, it has been suggested that the results may be accurate since sources of bias act in opposing directions.

3.4 Discussion: Comparison of CV and TTO

3.4.1 Feasibility

The majority of evidence seems to suggest that both methods are feasible for respondents to complete. Of the CV formats OE questions seem to elicit the lowest completion rates. Non-trading occurs in both methods, although more frequently in the TTO method. Non-trading in the CV method typically seems to be a true 'protest' zeros (e.g. the government should pay for this) rather than a true statement of preference. In contrast, in the TTO method non-trading is often a true preference for length of life, particularly when valuing own health or moderate health states as there may be a 'threshold' of tolerability that must be exceeded to induce a trading of time. Three studies were identified that allow direct comparison between TTO and CV (Krabbe *et al.* 1997; Jacobs *et al.* 2002; Kontodimopoulos and Niakas, 2006). Jacobs *et al.* (2002) ask 181 members of the general population to value a hepatitis A state though both TTO and WTP. In the WTP exercise 8% of respondents were unwilling to pay any money. In the TTO exercise between 10% and 25% (depending on state) were not prepared to pay any time (the exact proportion is not presented in the paper). Kontodimopoulos and Niakas (2006) asked 606 end stage renal disease patients to value their current health through TTO, and their WTP for an alleviation of their condition use a DC method with OE follow-up. The proportion of non-traders was very similar with both methods. In the TTO exercise 47.3% of respondents were not willing to give up any time to improve their health, and in the WTP exercise 40.5% of patients were unwilling to pay any money for treatment.

Qualitative feedback on the ease of completion seems to be favourable for both methods. While the TTO method has been shown to be feasible in self-complete format, it has been shown to perform better with the use of an interviewer setting with visual aids (Dolan et al, 1996b). In contrast, depending on the particular elicitation format used the CV method is perhaps easier to administer in a self-complete format.

3.4.2 Reliability

Different studies produce widely varying estimates of the test-retest reliability coefficient for both methods making it difficult to make comparisons. Furthermore, the figure produced will depend on the time that elapses between the first and second valuations. Broadly, both methods seem fairly reliable, frequently (but not always) producing correlation coefficients >0.7 which is suggestive of reliable performance (although as mentioned previously such an assessment is somewhat subjective).

Test-retest reliability measures can be biased if respondents are able to remember the value they gave in the first instance, which they then simply repeat in the second valuation. This is obviously more of a concern the shorter the time gap between the two valuations. It is possible that either TTO or CV could be more susceptible to such a bias. For example, perhaps respondents can more easily remember a monetary amount they specified than the number of years they traded. If this were the case, this bias would favour the CV method in relation to TTO.

While poor repeatability of valuations at the individual level poses questions over the validity of a method, it could be argued that what matters is stability at the aggregate level. From an economic evaluation perspective only aggregate values are utilised, so if these are consistent then individual values are of less importance. However, despite this studies tend to focus on correlation coefficients based on individual level variation and often do not perform simple tests of differences in means. Where these have been presented the results are mixed.

3.4.3 Theoretical Validity

CV has its theoretical roots firmly in welfarist theory, and WTP can be considered a measure of compensating or equivalent variation. Testing the how well the results comply with theory gives mixed results. The method has often been shown to be insensitive to scale, but results are usually positively correlated with ability to pay as would be expected.

While TTO has not emerged from a particular theory work has shown how the method can be considered to fit with welfarist theory, the standard TTO valuation being akin to the compensating variation for a gain. SG is typically seen as the gold standard, but it has been shown that SG and TTO utilities do not converge. Some may consider this concerning given the strong theoretical roots of the SG method. The standard explanation is that TTO utilities are biased downwards due to the concave nature of the utility function over duration. However, others have argued that SG utilities may also be susceptible to bias and the TTO method may in fact produce the more consistent results (as determined through comparison with direct ranking). Neither does the TTO method converge with VAS valuations, probably due to the effect of the 'threshold of tolerability' that must be exceeded in TTO valuations. However, the VAS method is generally considered weak due to the lack of a true trade-off, so the performance of TTO should not be judged in comparison to this measure.

TTO valuations generally do not suffer from the insensitivity to scale issues that are experienced with CV. However, the frequency of non-trading is suggestive of a degree of insensitivity for mild states.

3.4.4 Discriminant Validity

Most of the potential sources of bias in CV studies are a product of the study design, and while one source of bias can eliminated by amending the study design, it is likely that a new one will be introduced. For example, using a DC format may lead to yeah-saying bias, and move to a payment card method may eliminate this but at the expense of introducing range bias. In contrast, the difficulty with the TTO method is in the assumptions that need to be made to translate responses into utility values. This is reliant upon the underlying shape of individuals' indifference curves, which is difficult to measure and account for. The TTO method relies on a number of assumptions made by the QALY model, which are not necessary in the CV method e.g. constant proportional time trade-off. Both methods are potentially susceptible to loss aversion, meaning that they produce underestimates. It seems the difficulty with the CV method is in selecting an appropriate study design to elicit valid and consistent responses. The difficulty with the TTO responses is in turning them into valid utility values.

3.4.5 Direct Comparison

The three studies enabling direct comparison of the CV and TTO methods give mixed results on the convergence of the two methods. Kontodimopoulos and Niakas (2006) found a significant negative relationship between the two methods (the worse the heath state is perceived to be the more a respondent is willing to pay), while the other two studies found no significant relationship between them (Krabbe *et al.* 1997; Jacobs *et al.* 2002). The lack of a statistical relationship in these two studies may suggest that they measure independent aspects of preference.

The article by Smith (2001) also allows for direct comparison between the CV and TTO methods in terms of their sensitivity to changes in health status. They ask 50 subjects to values health states specified by the Assessment of Quality of Life instrument (Hawthorne *et al.* 1999) using both methods. Smith finds that WTP is more sensitive than TTO in distinguishing between different dimensions of health at the same nominal level of health status. However, this is based on a small sample size, with four WTP values not being significantly different, and eight TTO values not being significantly different. The study also finds that WTP was more sensitive to differences in quality of life between different levels of health within each dimension. Again this is based on small differences, with the TTO method yielding three insignificant relationships and the WTP approach yielding no insignificant relationships.

3.5 Identifying criteria against which an exploratory method can be assessed

The following chapters explore a new method, based on TTO, for estimating an MVQ. This chapter is useful in providing the basis for identifying criteria against which the performance of this method can potentially be assessed.

Feasibility will be assessed through completion rates and the prevalence of non-trading behaviour. However, non-trades only suggest the method is infeasible if they are protest responses rather than true statements of preference. Analysis will be performed to try and determine whether or not non-trades are meaningful statements of preference.

A key assessment of reliability is through test-retest. The study design in the following chapters does not allow for this particular test (questions are only asked on one occasion). As an alternative, when respondents value multiple outcomes, that follow a logical ordering, a simple analysis can be performed to determine what proportion of respondents valuations follow this logical ordering (note strictly speaking this is a test of logical consistency rather than reliability). This is only applicable in the main UK study in Chapter 6 as in the study in Chapter 4 respondents only value one outcome for each question.

Given that the method in the following chapters involves a trade off between length of life and income, similar validity criteria to those applied to CV can be used. Validity will be assessed by correlation with ability to pay and by sensitivity to scale. In addition, convergent validity will be assessed through convergence with a ranking exercise that precedes the main exercise (note, this is only performed in chapter 6 because no appropriate ranking exercise was included in Chapter 4).

3.6 Conclusion

Both the CV and TTO methods have uncertainties surrounding them. For the CV method there is a lack of clarity over the optimal way to design studies, and vulnerability of results to changes in study design lead to large variance between studies and hence reduced confidence in any results generated. There are still some question marks over whether the key assumptions required in the TTO method hold, and hence whether the results are meaningful.

Given the lack of consistency in MVQ estimates generated through the CV method, and the increased interest in the threshold in the UK with the proposed move to value based pricing, it may be timely to explore other methods. Of the preference based measures (VAS, SG and TTO) TTO may represent the most promising as it may produce more consistent results than the SG method, and is also easier to use. The next chapter will explore how the TTO method can be adapted to estimate an MVQ and present a pilot study.

CHAPTER FOUR

Exploring A New Method for estimating the Monetary Value of a QALY: initial testing (survey 1)

4.1 Introduction

In light of the need for greater information on how members of the general population value the QALY, this chapter seeks to test the feasibility of an entirely new method of eliciting an MVQ. The approach uses techniques that are an adaptation of the standard TTO exercise, as the questions ask respondents to trade off length of life to either achieve an income gain or avoid an income loss. This chapter outlines how information from these questions can be used to derive an MVQ. The study uses a sample of the Dutch general population, but the aim is not to derive a definitive MVQ for the Netherlands, but rather to determine the feasibility of this new method, to see if it can potentially overcome some of the problems encountered by the CV method, provide some initial evidence on its validity and to suggest improvements to the methods for the subsequent UK study. In line with the criteria for assessing a preference elicitation method outlined in the previous chapter, the new method will be assessed in terms of its feasibility and validity. Feasibility will be determined through an assessment of non-trading behaviour, while validity will be assessed by testing sensitivity to scale and correlation with ability to pay.

The following section outlines the theoretical basis and assumptions of the methods. The survey design and hypotheses are then presented. The results are then presented, before the implications for further work are discussed and conclusions drawn.

4.2 Methods

4.2.1 Theoretical Basis of the Method

TTO is a widely used choice-based method of health state preference elicitation with an inherent sense of sacrifice. It has been shown to be easier to complete than SG, and Buckingham and Devlin (2006) have outlined how the TTO method can be interpreted in the theoretical context of Hicksian Utility theory and hence comply with Welfare Economic principles in a similar fashion to WTP derived through CV. To recap, Buckingham and Devlin (2006) show that the standard TTO question elicits the compensating variation for a health gain. A respondent experiences a gain in health, and then accepts a reduction in life years to return them to their initial level of utility (see Figure 3.1 in Chapter 3 for the graphical presentation).

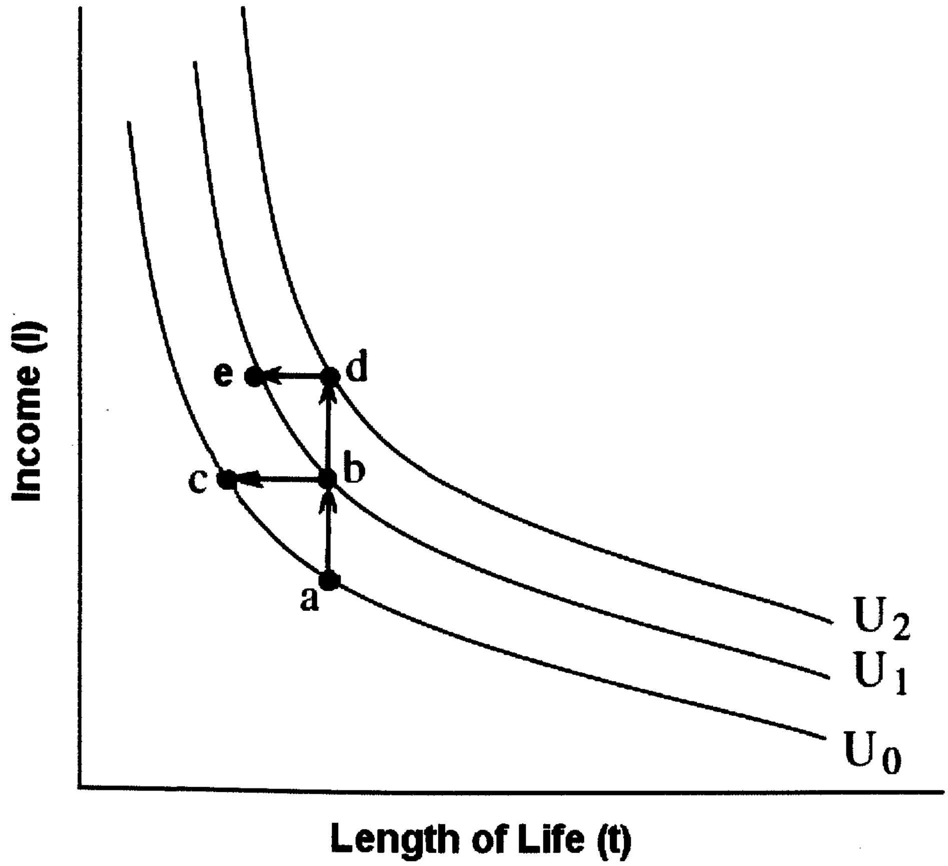
This chapter pilots two TTO questions both of which are based on the compensating variation for a gain, in which health is replaced by income so the trade off becomes between longevity and income rather than longevity and health. The difference between the two questions is in the indifference curve that they start on. The first question asks respondents the decrease in longevity that would be required to compensate for an increase in income. In the first question the increase in income is from an amount less than their current income, to their current income. The second question also asks respondents the decrease in longevity that would be required to compensate for an increase in income. However, in these questions the increase in income is from current income to a value higher than current income.

The wording of the first question looks as follows:

TTO 1: Trading years to avoid an ***income loss*** *in perfect health (<u>Compensating Income Loss</u>)*
"You can live for 10 years in perfect health with (100 - Y)% of your current annual income for each year and then die or you can live for a shorter period of time in perfect health with your current annual income for each year and then die." The length of time with current annual income is varied to reach a point of indifference.

The indifference curves representing the trade off are shown in Figure 4.1. The x axis represents length of life and the y axis represents income. Each indifference curve represents a level of utility that can be achieved by different combinations of longevity and income, where $U_2>U_1>U_0$. The first option is point *a* on indifference curve U_0 (10 years in perfect health with less than current income). The increase in income (to current income) moves the individual to point *b* on a higher indifference curve U_1. The respondent must then specify a decrease in longevity that returns them to their original indifference curve U_0, at point *c*.

Figure 4.1 Compensating Income Loss and Compensating Income Gain (adapted from Buckingham and Devlin, 2006, p.1151).



The second question also asks respondents the decrease in longevity that would be required to compensate for an increase in income, but the reference point differs:

TTO2: Trading years to achieve an ***income gain*** *in perfect health (Compensating Income Gain)*

"You can live for 10 years in perfect health with your current annual income for each year and then die or you can live for a shorter period of time in perfect health with (100 + Y)% of your current annual income for each year and then die." The length of time with current annual income is varied to reach a point of indifference.

Referring again to figure 4.1, the first option is at point *b* on indifference curve U_1 (10 years with current annual income). Note, in TTO2 the first option is on a higher indifference curve (U_1) than in TTO1 (U_0), because income is set at current annual income. An increase in income (to a value greater than current income) takes the individual on to a higher indifference curve U_2, at point d. The respondent must then specify a decrease in longevity that returns them to their original indifference curve at point *e* on U_1.

While both questions are compensating variations for a gain, TTO1 has been named Income Loss in order to distinguish between the two questions, and because the trade is actually to avoid an income loss rather than achieve an income gain. TTO2 will henceforth be referred to as Income Gain.

To see how the results from these questions can be used to derive an MVQ imagine that a respondent, facing a 20% income loss in TTO1 states that 9 years with normal annual income of €100,000 is equivalent to 10 years with 80% of this income, so €80,000. Therefore, using prospective lifetime income values (i.e. total income over the given duration) and an additive utility function (see 4.2.1.1 below) this point of indifference gives us the following information:

10U (Perfect Health) + €800,000 = 9U (Perfect Health) + €900,000 (1)

10U (Perfect Health) – 9U (Perfect health) = €900,000 - €800,000 (2)

U (Perfect Health) = **€100,000** (3)

The Income Gain data from TTO2 is analysed in a similar fashion to the Income Loss data in TTO1. Consider a respondent who is indifferent between 10 years with their current income and 9 years with 120% of their current income (9*120%=1080%). Their income is, once again, €100,000 per year:

10 U(PH) + €1,000,000 = 9 U(PH) + €1,080,000 (4)

10U (PH) – 9U (PH) = €1080,000 – €1,000,000 (5)

U(PH) = **€80,000** (6)

Some explanation of the effect of the different inputs on the MVQ is warranted. The more years that are traded the smaller the MVQ value. If the same number of years is traded in the Income Loss and Income Gain questions, the Income Loss results will be higher. If the income change level doubles in the Income Loss questions, a doubling of the number of years traded will lead to the same results (providing income remains constant). In the Income Gain questions if the income change level doubles and the number of years traded also doubles, the MVQ value will less than double. Changes in income have a proportional effect on the MVQ i.e. if income is doubled, the MVQ will also double. Note that these effects occur at the individual level and so may not occur perfectly in the results tables which present summaries at the aggregate level.

4.2.1.1 Assumptions of the methods

The above calculation methods rely on a number of assumptions:

1) **Additive Separability between health and income in the utility function.**

The method assumes that the utility derived from a given duration is not determined by the level of income. In reality a year in perfect health may be valued higher when combined with a higher amount of income. If this additive separability assumption does not hold the results are likely to be overestimates. When faced with fewer years and higher income, each of these years will be valued more highly than in the initial scenario of more years with lower income. Therefore fewer years will be needed to achieve the compensating utility increment. The result of fewer years being traded is to increase the MVQ estimate.

2) **Constant marginal rate of substitution between health and income.**

Referring to figure 4.1 above the method assumes that the indifference curves are straight lines, and is trying to estimate the slope of that line. The marginal rate of substitution is equivalent to ratio of the two diminishing marginal utilities. So, by assuming a constant rate of substitution between health and income the method assumes that marginal utilities for health and income diminish at equivalent rates. Relaxing this assumption would require us to estimate an indifference curve across a range of combinations of health and income, to identify the shape of the relationship between the two. Unfortunately, the sample size generates a dataset that does not cover a wide enough range of values to make this possible.

3) **No Time Preference**

This is a standard assumption made in TTO studies, although evidence suggests this results in a small downward bias in results (Attema and Brouwer, 2009). This is not a fundamental assumption: if individual time preference for health and income were obtained they could be incorporated into the calculation of an MVQ.

Note, the interpretation of the results of the above calculations relies on a key assumption of the QALY model: quality of life and length of life are commensurate and interchangeable. The method involves no Quality of Life valuation, so it could be argued that the results represent the monetary value of a life year rather than a QALY. However, other studies have elicited the WTP for a life extension and presented their results as an MVQ (see for example Shiroiwa *et al.* 2010). This issue will be addressed further in the discussion section.

4.2.1.2 Lexicographic responses

If respondents do not trade any years in the exercises, calculating an MVQ for them becomes problematic because the left hand side of equation (2) becomes 0, meaning that the equation would give an indeterminate value for these respondents. If such responses occur and are a protest against the exercise this poses questions over the feasibility of the exercise. If such responses are a meaningful statement of preference for a seemingly infinite preference for life over income then this does not mean the exercises are infeasible, but means the calculation method above is not capable of calculating a finite MVQ for such individuals based on these meaningful responses. These lexicographic (henceforth LG) responses can be displayed using indifference curve analysis as in Figure 4.2.

Figure 4.2 – Indifference curve showing infinite preference (lexicographic) for length of life over income

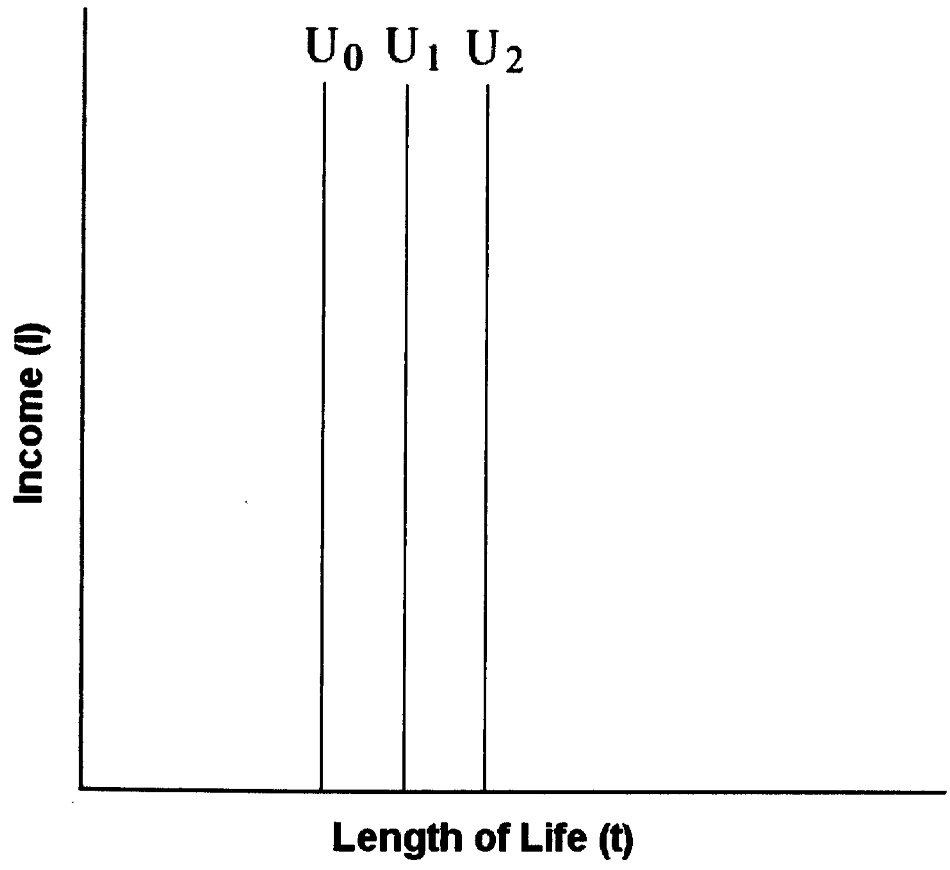


Figure 4.2 shows vertical indifference curves representing an infinite preference for length of life over income. Since a given indifference curve shows the combinations of length of life and income that deliver a given level of utility, vertical curves show that a given length of life delivers a certain amount of utility regardless of income. Therefore, a respondent with lexicographic preferences of this nature would not give up any length of life to increase their income. In the context of the Income Loss questions, the increase in income that the respondent is faced with (from less than current income to current income) does not increase their utility, therefore no decrease in longevity is required to return them to their initial level of utility.

It should be noted that non-trading in the Income Loss or Income Gain questions does not necessarily mean that the indifference curve is perfectly vertical as in Figure 4.2 above, it just means that the curve is sufficiently steep that the utility gained from the increase in income is less than the amount of utility that would be lost through giving up the smallest amount of longevity possible (the smallest unit of trade was one month). Furthermore, non-trading for a given income change level does not mean that the *entire* indifference curve is vertical (or sufficiently steep) as in Figure 4.2, it only determines the slope of the indifference curve between the two income points on the y axis that the respondent is being questioned on.

Whether non-trades are protest responses or a true reflection of LG preferences, if an individual calculation method (i.e. calculate an MVQ for each individual and then aggregate) is to be used, then non-traders must be excluded. An alternative is to use an aggregate approach (i.e. use aggregate income and aggregate number of years traded) but this represents a movement away from standard welfare economics (societal welfare as the sum of individual welfare). Results are presented from both approaches. Throughout the remainder of the thesis responses in which individuals will not give up any time to increase their income are referred to as LG responses. This is so that when two new questions are introduced in the following chapter the same terminology can be applied to different responses that have the same interpretation, (in these new questions LG responses occur when respondents give up all of their income, hence non-trading would not be appropriate terminology).

4.2.1.3 Over-traders

For TTO1 if the percentage of life years the respondent is prepared to give up is larger than the percentage loss in annual income he is faced with then his MVQ will be negative. In other words if the respondent is faced with a 20% loss in annual income, and if they trade more than 2 years of life their MVQ value will be negative (since total income is negative in the right hand side of equation 2). If they trade exactly 2 years their MVQ value will be zero. So for a 40% loss they cannot trade more than 4 years, and for a 60% loss they cannot trade more than 6 years. For TTO2 the relationship is not linear. For a 20% gain they cannot trade more than 1.666 years (=[20/120] * 10), for a 40% gain they cannot trade more than 2.86

years (=[40/140] * 10) and for a 60% gain they cannot trade more than 3.75 years (=[60/160] * 10).

If these responses are arising because respondents do not understand the exercise this poses questions over the feasibility of the questions. However, if they are meaningful statements of preference this suggests a weakness in the calculation method. In the individual approach negative MVQ values are truncated at zero. In the aggregate approach the number of years traded is left unchanged.

4.2.2 General Design

The survey had two aims: the first was to determine whether or not respondents valuing hypothetical health states through TTO included income effects in their valuations; the second was to explore the feasibility of a TTO-based MVQ. Five different TTO questions (including the standard MVH TTO) were asked. The first three questions addressed the first aim and are not presented in this thesis, but further details can be found in Tilling *et al.* (2012).

The final two TTO questions of the survey are of relevance to this chapter and correspond to TTO1 (Income Loss) and TTO2 (Income Gain) outlined above.

Respondents were randomised to one of three income change levels (Y), either 20%, 40% or 60% (variants A, B, and C of the questionnaire respectively), which they then faced in both TTO1 and TTO2. Since the survey was administered in an online self-complete fashion there was no iterative process. Respondents were simply asked to state how many years with higher income, was equivalent to 10 years with lower income. There was no visual aid. The smallest unit of trade was one month. It should also be noted that all respondents received the two questions in the same order: TTO1 followed by TTO2. Therefore, we cannot rule out the possibility that responses to TTO2 are affected by respondent's having already seen and given an answer to TTO1.

The three TTO questions not presented here involved valuing 4 EQ-5D health states each. The two TTO questions used to derive an MVQ involved one valuation each. Therefore, there were a total of 14 TTO questions asked during the survey.

4.2.2.1 Background Characteristics

Background questions were asked of respondents at the start of the exercise. The characteristics and response options given are outlined in Table 4.1.

Table 4.1 Background characteristics collected at the start of the interview

Background Characteristic	Categories
Gender	Male/Female
Age	Open-ended
Education	Left school at 12/ vocational study to 16/ academic study to 16/ academic study to 18/ vocational study to 20/ have a degree/ have post-graduate qualifications
Marital Status	Single/ Married or Co-habiting/ divorced/ widow or widower
Children	Yes/No, if yes then how many and what ages
Occupation	In paid employment – if so temporary or permanent contract, and for how many hours per week/ self-employed – if so how many hours per week do you work on average/ housewife/husband/ retired/ incapable of work/ student
Nationality	Dutch/ Turkish/ Surinamese/ Moroccan/ Antillian/ Other
Religion	Protestant/ Roman Catholic/ Muslim/ Jewish/ Hindu/ Other/ Not religious
Net Own Monthly Income	Chosen from six income brackets ranging from 'no income' to more than €2,750 per month
Net partners monthly income	Chosen from same six income brackets as above. One further option included for 'I do not have a partner'

Respondents were also asked to report their own health on the EQ-5D and value it on the VAS as part of the background questions. The VAS scale ranged from best imaginable health (100) to worst imaginable health (0). Respondents also placed 'dead' and EQ-5D state 11111 on the VAS scale.

Predicting correlations between demographics and responses is difficult given the exploratory nature of the exercise. Previous TTO studies have found that women give values that are lower then those given by men (Dolan *et al.* 1996b) i.e. they trade more years. This might suggest that they attach less value to longevity (alternatively they attach more value to quality of life). If this is the case then they might trade more years in the TTO exercises in this study. Dolan *et al.* (1996b) also found an inverse U shaped age relationship in TTO responses, whereby valuations increased slowly from the age of 18 to about 40, then began to fall slowly from about 40 to 60 before falling sharply in later years. It could be that older respondents are less prepared to tolerate suffering in later life, in which case the relationship would not be expected to hold in this study as the questions specify perfect health. However, it could be that elderly respondents who do not believe they have 10 years left to live are more prepared to give up these excess years. The study only recruits respondents of working age (18-65) so again we would not expect this relationship to hold.

It is anticipated that married respondents will trade fewer years (more likely to give LG responses) because they are able to rely on their partners income. Respondents with children might trade fewer years because they want to be around to look after their children. Respondents in employment might have savings and hence be prepared to trade fewer years. Religion was included as this may offer an explanation for respondents who refuse to trade any time to increase their income. Income is obviously a key input into the calculation method. Predicting the effect of income on trading behaviour is difficult. Someone with higher income is more able to manage a reduction in their income, but they also will lose more income following a percentage reduction. It is anticipated that someone who's partner has a high income will trade fewer years because they could rely on their partners income following a reduction in their personal income.

More educated respondents might be less likely to over-trade if they have a better understanding of the task and are able to identify the point at which total lifetime income becomes lower in the alternative scenario (i.e. in the right hand side of equation 2 from earlier).

4.2.3 Analysis and Null Hypotheses

4.2.3.1 Background Characteristics and Extreme Non-Traders

Some respondents did not trade in any of the 14 TTO exercises. It was felt that this was due to the nature of an online survey, with respondents simply clicking through the exercise in order to finish as quickly as possible. These respondents are excluded from the entire analysis. Background characteristics are presented for these "extreme non-traders" and the rest of the sample, with Chi^2 tests determining if any differences are significant. Background characteristics are also presented by questionnaire variant, with Chi^2 tests again determining any significant differences.

4.2.3.2 Feasibility

Feasibility is tested by considering the percentage of respondents that either give LG responses or over-trade. The number of LG responses and over-traders are presented for each of the two TTO questions and for each level of income change. A respondent could potentially be an LG responder in one of the two questions they face, and an over-trader in the other question.

To try and determine whether these responses are true statements of preference, or whether they are protest responses (in the case of LG responses) or caused by poor understanding of the exercise (in the case of over-trades), analysis is performed to identify any determinants of being an LG responder or over-trader.

Chi^2 tests are used to compare background characteristics by LG responders and the rest of the sample, for the Income Loss and Income Gain questions. This analysis pools respondents from the three versions of the questionnaire. Logit regressions are used to further test the effect of background characteristics on the likelihood of being an LG responder. Separate regressions are used for the Income Loss and Income Gain questions. Four regressions are performed for each question. One regression is performed for each of the income change levels, and a further model pools all three income change levels. The binary dependent variables are whether or not a respondent gave an LG response, and the explanatory variables are the background characteristics. In the pooled models dummies are included to show the effect of higher income change levels. The Hosmer-Lemeshow test is used as an indication of model fit (Hosmer and Lemeshow, 2000). This test groups observations into deciles using the percentiles of the estimated probabilities. It then compares the expected and observed number of successes in each of these deciles, using a Chi^2 distribution. Since the Pseudo R^2 is only a weak indication of model explanatory power, the Akaike Information Criterion is presented to show explanatory power and allow comparison between models. Akaike's (1973) information criterion is defined as:

$$AIC = \{-2\ln L(M_k) + 2(P_k)\}/N$$

where $L(M_k)$ is the likelihood of the model and P_k is the number of parameters in the model. All else being equal, the model with the smaller AIC is considered the best fitting model (Long and Freese, 2006).

Further probit regressions, including the same explanatory variables as the models described above, are performed to test the effect of background characteristics on the likelihood of being an over-trader in both the Income Gain and Income Loss questions. Only one model, pooled across the three income change levels, is used for each of the two questions (due to the small number of over-traders).

4.2.3.3 Validity

Internal Validity is tested in two ways, through sensitivity to scale and correlation with ability to pay.

4.2.3.3.i Sensitivity to Scale

Sensitivity to Scale is considered in two ways. Firstly, the sensitivity to the income change level is considered. Sensitivity of both the number of years traded by respondents and the sensitivity of the MVQ estimates produced by these trades is studied. The mean numbers of years traded, both including and excluding LG responders, are presented for each of the question types and each of the income change levels. The first two null hypotheses are as follows:

Null Hypothesis 1: The number of years traded in the Income Loss questions will be the same regardless of the level of income loss

Alternatively, we would expect respondents to trade more years to avoid a larger income loss. Given diminishing marginal utility of income we might expect respondents to trade years at an increasing rate as the level of income loss increases. However, there is also likely to be diminishing marginal utility of life years, so the exact rate at which the trade of years increases will be determined by the relative diminishing marginal utilities of income and life years.

Null Hypothesis 2: The number of years traded in the Income Gain questions will be the same regardless of the level of income gain.

Alternatively, we would expect respondents to trade more years to achieve a larger income gain. Given diminishing marginal utility of both income and life years we would expect the number of years traded to increase at a decreasing rate as the level of income gain increases.

The second way sensitivity to scale is tested is by comparing the number of years traded in the Income Loss and Income Gain questions. Although this is not a typical test of sensitivity to scale, given diminishing marginal utility of income, as mentioned in the above hypotheses we would expect differences between the two questions. This gives rise to a third hypothesis:

Null Hypothesis 3: The number of years traded will not differ between the Income Loss and Income Gain questions.

Alternatively, given diminishing marginal utility of both income and life years, we would expect respondents to trade more years in the Income Loss questions. This is also supported by the findings of Kahneman and Tversky (1979): through a series of probabilistic choices they found risk aversion in choices involving sure gains, and risk seeking involving sure losses.

The above three hypotheses are tested through t-tests. Unpaired t-tests are used to compare the number of years traded by income change level, while paired t-tests are used to compare the number of years traded by type of question. These two tests of sensitivity to scale are also performed for the MVQ estimates. See section 4.2.3.4 below for an explanation of the MVQ values that are presented. Testing the number of years traded is a purer test of sensitivity to scale as this is the exact behaviour of the respondent, while the MVQ estimates are derived through the calculation method outlined earlier in section 4.2.1. The nature of the calculation method means we would expect the MVQ estimates to be less sensitive to scale than the number of years traded. For example, in the Income Loss questions if the number of years traded increased proportionately with the level of income loss the MVQ estimate would remain constant. In the Income Gain questions if the number of years traded increases at a decreasing rate the MVQ estimates will increase.

4.2.3.3.ii Correlation with ability to pay

The second test of validity is that the MVQ estimates should be influenced by ability to pay. This gives rise to the fourth null hypothesis:

Null Hypothesis 4: *The MVQ results will not differ depending on respondent income.*

Alternatively, given the importance of respondent income in the calculation method (see equations 1-6) for both questions, we would expect the MVQ estimates to be positively correlated with respondent income.

To test this hypothesis MVQ results are presented by respondent income bracket. The null hypothesis is tested by observing any trends in these results. Statistical testing is not feasible due to the small sample sizes in the different income brackets.

4.2.3.4 MVQ Results

MVQ results are first presented for the **individual approach**, excluding LG responders, based on both individual income and household income. Negative values (caused by over-traders) are truncated at zero. Mean MVQ results (based on individual income) are compared through both paired and unpaired t-tests. Further MVQ estimates based on different sub-samples (with a balanced panel, and excluding over-traders) can be found in the Appendix.

MVQ results generated through the **aggregate approach** are also presented. Results are presented including LG responses and over-traders. The results are presented based on both individual income and household income. Further sets of aggregate results are presented in an Appendix: Excluding LG responders and including over-traders; excluding LG responders and over-traders; excluding LG responders and including over-traders, with a balanced panel.

4.2.4 Respondent Income

In order to determine the level of "current annual income" for each respondent, respondents were asked to choose the income bracket within which their monthly income fell in the background characteristics questions. For our analysis these income brackets were converted into numerical values using the mid-point of each bracket (Layard et al. 2008). For respondents in the lowest income bracket an income of two thirds of the upper limit of the bracket was used. For respondents in the highest income bracket an income of 1.5 of the lower income limit of the bracket was assumed (Layard et al. 2008).

Respondents were also asked to specify their partner's income by choosing from the same set of income brackets. Household income is calculated by simply adding an individual's income with their partner's income.

4.2.5 The Data

Data were gathered through an online self-completion questionnaire that was in Dutch and was administered in the Netherlands. Invitations were sent out to a subset of an existing panel of potential survey respondents in order to obtain a representative sample of 300 members of the Dutch general public. We selected respondents between the ages of 18 and 65 as we felt that questions about income were most relevant for people in this age bracket. The data collection was performed by an online market research company (Survey Sampling International; www.surveysampling.com).

4.3 Results

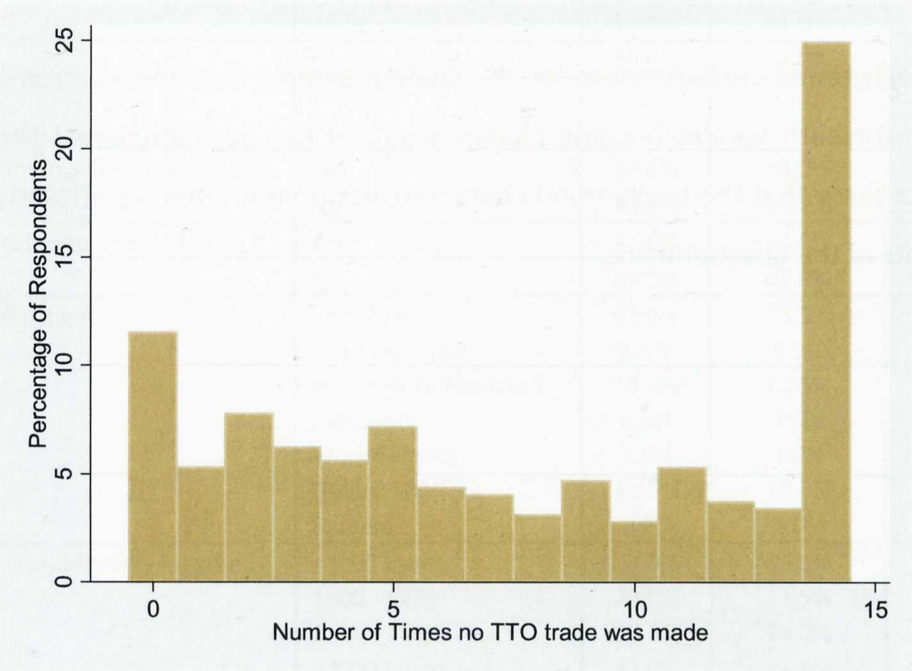
This section first presents background characteristics by traders and 'extreme non-traders', and then by questionnaire variant. Following this LG responses and over-trading in the context of TTO1 and TTO2 are explored. Validity is then assessed through sensitivity to scale and correlation with ability to pay. Finally, the MVQ results are summarised.

4.3.1 Background Characteristics and 'Extreme Non-Traders'

Data are available from 321 members of the Dutch general public who participated in the online survey. Preliminary data examination showed that many respondents had been unwilling to trade any life years in a number of the 14 TTO exercises. Figure 4.3 illustrates the number of TTO exercises in which respondents were not prepared to trade time for improved health/income. This shows that 25% of respondents were unwilling to trade any time in any of the 14 TTO exercises. For some respondents this may be a genuine representation of preferences but we suspect that many of these respondents strategically chose not to trade. Respondents were selected from a database of individuals who have signed up to complete online exercises of this nature. Although the TTO protocol was not iterative respondents may have non-traded in order to avoid having to think about their response, and hence complete the exercise more quickly. The sooner they complete the exercise the sooner they are awarded a given amount of money to be donated to a charity of their choice and the chance to win a prize themselves. Van Nooten et al. (2009) also found numerous respondents opted not to trade in TTO exercises in their online questionnaire. A further five respondents traded the same number of years in all 14 TTO exercises.

Table 4.2 shows the background characteristics firstly for the entire sample and then for those that have traded in at least one of the TTOs and those that have not traded at all (i.e. 'extreme' non-traders). The full sample has slightly more males than females. Forty two percent of the sample were not employed which, given that income in the two TTO questions in this chapter is presented as a percentage, is likely to influence a respondents willingness to give up life years. More than half of the sample had children, which is also likely to affect a respondent's willingness to give up life years. Just under half of the sample are married and the mean VAS score for own health was 0.76 (where 1 represents full health and 0 represents being dead).

Figure 4.3 - Histogram showing the number of TTOs in which respondents were unwilling to trade



There were significant differences between the two groups for three variables. Extreme non-traders were more likely to be parents than traders. Extreme non-traders were also more likely to be married. Self-reported health on the VAS was weakly correlated with whether or not respondents traded, with non-traders being in better health than traders.

The likely strategic responses of extreme non-traders may dilute the more meaningful responses of traders and will not help us effectively assess the performance of the two new measures. We have therefore chosen to exclude these extreme non-traders from our analysis which reduces the sample size from 321 to 241. Note, respondents may have non-traded in the Income Loss and Income gain questions but traded in one or more of the other TTO questions, in which case they are still included in the analysis. A total of 41 respondents gave negative VAS valuations of own health (13 of whom were extreme non-traders). It is very unlikely that someone in a state of health worse than dead would be able to complete an online questionnaire. Examination of these responses suggested that they were not meaningful, and were predominantly caused by very high valuations of dead. Comparison with their self reported health in EQ-5D showed that these respondents were generally in good health (modal EQ-5D profile for this group was 11111). These respondents with negative VAS for own health were excluded from analysis involving VAS of own health (reducing sample size to 213), but included in all other analyses.

Table 4.3 shows the background characteristics for the analysis sample (excluding extreme non-traders) and then for each variant (income change level) of the questionnaire. The results of the Chi^2 tests show that the background characteristics do not differ significantly across the three variants of the questionnaire.

Table 4.2 - Background characteristics by traders and extreme non-traders

		All	Traders	Extreme Non-Traders	Chi[2] Test[1] (p-values)
Number of Respondents		321	241	80	
Gender	Male	51.0%	52.0%	54.0%	0.350
	Female	49.0%	48.0%	46.0%	
Age	Average (SD)	44(13.1)	43.19 (13.19)	46.6 (12.37)	0.148
	18-35	29.0%	32.0%	21.0%	
	36-50	32.0%	31.0%	33.0%	
	51-65	39.0%	37.0%	46.0%	
Educated beyond the minimum school leaving age	Yes	67.0%	66.0%	70.0%	0.507
	No	33.0%	34.0%	30.0%	
Educated to Degree Level	Yes	31.0%	32.0%	29.0%	0.592
	No	69.0%	68.0%	71.0%	
Employment	Employed	52.5%	53.5%	50.0%	0.874[2]
	Self-Employed	5.5%	5.0%	7.5%	
	House Wife/Husband	13.0%	12.5%	15.0%	
	Pensioner	6.5%	7.0%	5.0%	
	Work Seeking	3.0%	3.0%	2.5%	
	Unable to Work	11.5%	10.0%	16.0%	
	Student	8.0%	9.0%	4.0%	
Net Own Monthly Income	<1000 Euros	39.0%	38.0%	41.0%	0.873
	1000 - 1499	22.0%	21.5%	24.0%	
	1500 - 1999	18.0%	19.0%	16.0%	
	>2000 Euros	21.0%	21.5%	18.0%	
Children	Yes	54.0%	49.5%	67.5%	**0.005**
	No	46.0%	50.5%	32.5%	
Religion	Protestant	17.0%	16.5%	19.0%	0.182
	Roman Catholic	26.5%	28.5%	20.0%	
	Atheist	49.5%	49.5%	50.0%	
	Other	7.0%	5.5%	11.0%	
Marital Status	Married	46.5%	42.5%	59.0%	**0.013**[3]
	Single/Never Married	21.0%	22.5%	16.0%	
	Divorced	10.0%	12.0%	4.0%	
	Widowed	2.0%	2.0%	1.0%	
	Living Together	17.5%	18.0%	17.5%	
	Other	3.0%	3.0%	2.5%	
Mean Self-Reported Health on the EQ-VAS[4]		0.76	0.75	0.80	**0.073**[5]

1. Test between traders and extreme non-traders
2. Employment treated as dichotomous, employed/self-employed vs the rest.
3. Marital status treated as dichotomous, married vs the rest.
4. Sample Size: All (280), Traders (213), Extreme Non-Traders (67)
5. This is a t-test

Table 4.3 - Background Characteristics by questionnaire variant

		Analysis Sample	Variant A (20% income change)	Variant B (40% income change)	Variant C (60% income change)	Chi^2 Test (p-values)
Number of Respondents		241	78	80	83	
Gender	Male	52%	50%	56%	49%	0.683
	Female	48%	50%	44%	51%	
Age	Average	43.19	43.71	42.91	42.96	0.808
	SD	13.19	12.96	13.20	13.52	
	18-35	32%	32%	31%	33%	
	36-50	32%	32%	28%	35%	
	51-65	36%	36%	41%	32%	
Educated beyond the minimum school leaving age	Yes	66%	69%	64%	65%	0.750
	No	34%	31%	36%	35%	
Educated to Degree Level	Yes	32%	37%	30%	29%	0.479
	No	68%	63%	70%	71%	
Employment	Employed	54%	58%	45%	58%	0.701[1]
	Self-Employed	5%	4%	10%	1%	
	House Wife/Husband	12%	10%	18%	10%	
	Pensioner	7%	9%	4%	8%	
	Work Seeking	3%	5%	1%	2%	
	Unable to Work	10%	10%	11%	9%	
	Student	9%	4%	11%	12%	
Net Own Monthly Income	<1000 Euros	38.0%	37.0%	46.0%	31.5%	0.237
	1000 - 1499	21.5%	18.0%	20.0%	26.5%	
	1500 - 1999	19.0%	20.5%	20.0%	15.5%	
	>2000 Euros	21.5%	24.5%	14.0%	26.5%	
Children	Yes	49%	49%	52%	47%	0.773
	No	51%	51%	48%	53%	
Religion	Protestant	17%	14%	16%	19%	0.461
	Roman Catholic	29%	33%	28%	25%	
	Atheist	49%	44%	51%	3%	
	Other	5%	9%	5%	53%	
Marital Status	Married	43%	41%	47%	40%	0.567[2]
	Single/Never Married	22%	19%	19%	29%	
	Divorced	12%	10%	16%	9%	
	Widowed	2%	3%	3%	1%	
	Living Together	18%	24%	12%	17%	
	Other	3%	3%	3%	4%	
Mean Self-Reported Health on the EQ-VAS[3]		0.75	0.73	0.73	0.77	0.131[4]

1. Employment treated as dichotomous, employed/self-employed vs the rest.
2. Marital status treated as dichotomous, married vs the rest.
3. Sample Size: Analysis Sample (213), Variant A (69), Variant B (69), Variant C (75)
4. This is a t-test

4.3.2 Feasibility

Feasibility was assessed by the prevalence of LG responses and over-trading behaviour. Table 4.4 illustrates the scale of the problem posed by LG responders and over-traders. In the 20% Income Loss questions 72% of respondents did not trade any time to avoid the income loss. This value falls to 45% in the 60% *loss* questions. There are more LG responders in the gain questions for the two larger income change levels, which is consistent with the theory of diminishing marginal utility of income. While the proportion of LG responses in the Income Loss questions decreases as the level of loss increases, the proportion remains fairly constant for the Income Gain questions. The vast majority of people who do give LG responses in the Income Loss questions, also do so in the corresponding Income Gain questions, (e.g. Of the 49 respondents who do not trade in the 20% Income Gain question, 45 of them also non-trade in the 20% Income Loss question). A far smaller proportion of respondents over-trade than give LG responses (e.g. 72% of respondents give LG responses in the 20% Income Loss question, but only 14% over-trade in this question). Also, the behaviour of over-trading appears to be less consistent. There is no clear pattern across the income change levels for the Income Loss questions. For the Income Gain questions, the proportion of over-traders decreases as the level of income change increases. Approximately half of those respondents who over-trade in the Income Loss questions also over-trade in the Income Gain questions. The proportion of non-problematic responses is small, ranging from 14% for 20% loss, to 40% for 60% loss.

Table 4.4 - Number of LG responses and Over-Trades across the three income change levels

	Variant A: 20%			Variant B: 40%			Variant C: 60%		
	Loss	Loss *and* Gain	Gain	Loss	Loss *and* Gain	Gain	Loss	Loss *and* Gain	Gain
Number of Respondents	78	78	78	80	80	80	83	83	83
Number who gave LG responses	56	45	49	43	38	52	37	32	53
% LG responses	72%	58%	63%	54%	48%	65%	45%	39%	64%
Number who Over-Traded	11	7	17	16	8	14	13	7	14
% Over-Traders	14%	9%	22%	20%	10%	18%	16%	8%	17%
Non-problematic Responses	11	9	12	21	9	14	33	15	16
% Non-Problematic	14%	12%	15%	26%	11%	18%	40%	18%	19%

It is not possible to determine whether the LG responses are 'protest zeros' or true statements of preference for length of life over income. Given that respondents who non-traded in all 14 TTO exercises are already excluded, the results above are not influenced by respondents who non-trade simply to complete the exercise. However, it is perfectly feasible that respondents may find the earlier (standard) TTO exercises acceptable, but consider these particular questions involving a trade off between length of life and income unacceptable (Alternatively it could be that fatigue or even boredom sets in later in the questionnaire). Indeed, the level of non-trading (LG responses) was considerably higher in the income loss and income gain questions than in the TTO valuations that preceded them. At the end of the survey respondents were asked how difficult they found it to imagine being in a hypothetical health state, where 0 was not at all difficult and 10 was very difficult. This question corresponds to the 12 TTO questions that precede the Income Loss and Income Gain questions. They were also asked to consider how difficult they found it to take into account the impact of income. This question is relevant to all of the TTO questions. Interestingly respondents found it slightly harder to imagine being in a hypothetical state of health (mean 7.11) than they did to take account of income (mean 6.75).

The calculation method means that over-traders generate values that cannot be used, and it may seem illogical that respondents would want to trade to the extent that their lifetime income becomes lower. However, in the Income Loss questions this might be explained by diminishing marginal utility of income, while in the Income Gain questions respondents may want to live for a shorter period of time with high income, than a longer period of time with lower income, even if the lifetime income is lower.

Considering the relationship between background characteristics and LG responses/over-trading behaviour will help to determine whether the behaviour is meaningful and hence whether it represents a true statement of preference or an unsystematic protest response.

4.3.2.1 The relationship between background characteristics and LG responses

Background characteristics were compared by whether or not respondents gave LG responses through Chi^2 tests (see Appendix A.4.1 and A4.2). For the Income Loss questions four variables were statistically significantly different between the LG responders and the rest of the sample. More LG responders were educated beyond the minimum school leaving age (significant at the 10% level). A higher proportion of LG responders were in employment, and fewer non-traders were housewives (significant at 5% level). Finally, self-reported health on the VAS was higher among the LG responders (significant at 10% level). Comparing background characteristics by whether or not respondents were LG responders in the Income Gain questions did not find any significant differences.

Tables 4.5 and 4.6 show logit regressions used to assess the effect of background characteristics on the likelihood of respondents giving LG responses in both the Income Loss and the Income Gain questions. VAS of Own Health is not included in the individual models because this would have required exclusion of the inconsistent responses. In the pooled models it was felt the sample size was large enough to allow the inclusion of these responses. Although there is limited significance in the models they do suggest that in the 40% *loss* questions and 40% gain questions respondents in employment (or self-employed) are highly significantly (1%) more likely to give an LG response. There was a weakly significant U-shaped age relationship in the pooled loss model, the bottom of the curve occurring at age 44. There was also a weakly significant U-shaped age relationship in the 40% gain model. The pooled model for the Income Loss questions shows respondents were less likely to give LG responses when faced with a larger Income Loss.

Table 4.5 - The effect of background characteristics on the likelihood of being an LG responder in the Income Loss questions				
	Version 1: 20%	**Version 2: 40%**	**Version 3: 60%**	**Pooled across all three versions**
Variable	**(n=78) (Pseudo R^2=0.112)**	**(n=80) (Pseudo R^2=0.120)**	**(n=83) (Pseudo R^2=0.069)**	**(n=213) (Pseudo R^2=0.102)**
Income>999euros per month	0.166	-0.142	-0.115	-0.122
Gender (Male=1, Female=0)	-0.156	0.070	-0.089	0.024
Age	0.002	-0.025	-0.042	**-0.044***
Age^2	0.000	0.000	0.001	**0.001****
Married=1, Other=0	0.096	-0.065	-0.150	-0.037
Educated Beyond Minimum School Leaving Age	0.122	0.113	0.002	0.097
Have a Degree	0.148	-0.014	0.018	0.038
Working=1, Not Working=0	-0.093	**0.407*****	0.192	**0.285*****
Have Insurance	0.008	-0.175	0.111	-0.051
Have Children	0.109	0.117	-0.056	-0.002
Religious=1, Atheist=0	0.149	-0.043	0.087	0.090
VAS Own Health				-0.237
VAS Own $Health^2$				0.351
40% loss vs 20% loss				**-0.175***
60% loss vs 20% loss				**-0.289****
Akaike Information Criterion	1.364	1.515	1.569	1.383
Hosmer-Lemeshow's Test	0.170	0.118	0.158	0.199
Values presented are marginal effects coefficients. Significance is shown as follows: * 10%, **5%, ***1%				

Table 4.6 - The effect of background characteristics on the likelihood of being an LG responder in the Income Gain questions

	Version 1: 20%	**Version 2: 40%**	**Version 3: 60%**	**Pooled across all three versions**
Variable	**(n=78) (Pseudo R^2=0.068)**	**(n=80) (Pseudo R^2=0.108)**	**(n=83) (Pseudo R^2=0.025)**	**(n=213) (Pseudo R^2=0.061)**
Income>999euros per month	**0.267***	**-0.253***	0.122	-0.036
Gender (Male=1, Female=0)	0.092	0.044	0.047	0.100
Age	-0.004	**-0.065***	-0.028	-0.032
Age^2	-0.000	**0.001***	0.000	0.000
Married=1, Other=0	0.027	-0.081	-0.030	-0.002
Educated Beyond Minimum School Leaving Age	-0.048	0.005	-0.073	-0.046
Have a Degree	0.105	0.056	-0.030	0.093
Working=1, Not Working=0	-0.079	**0.333****	0.087	0.147
Have Insurance	-0.085	-0.271	0.043	-0.124
Have Children	0.185	0.191	0.080	0.087
Religious=1, Atheist=0	0.103	0.041	0.048	0.089
VAS Own Health				-1.026
VAS Own $Health^2$				**1.021***
40% gain vas 20% gain				0.042
60% gain vas 20% gain				0.040
Akaike Information Criterion	1.538	1.455	1.565	1.373
Hosmer-Lemeshow's Test	0.145	0.199	0.156	0.273

Values presented are marginal effects coefficients.
Significance is shown as follows: * 10%, **5%, ***1%

The Hosmer-Lemeshow test statistics in Tables 4.5 and 4.6 are not indicative of model misspecification. The Akaike Information criterion values suggest that the pooled models have the greatest explanatory power.

4.3.2.2 Is there an explanation for over-trading behaviour?

Background characteristics by whether or not respondents over-traded in both the Income Loss and Income Gain questions are shown in the Appendix (Tables A4.3 and A4.4). There are no significant differences between over-traders and the rest of the sample in either table.

Table 4.7 shows the results of logit regressions for the effect of background characteristics on the likelihood of being an over-trader. Due to the small number of over-traders it would not have been meaningful to present separate models for each version of the questionnaire, so we just present two models, pooled across the three versions. In the Income Gain questions no variables have a significant effect on being an over-trader. However, in the Income Loss questions Age and Age^2 have highly significant (1%) impacts. Age is positive and Age^2 is negative, suggesting an n-shaped age relationship i.e. as age increases the likelihood of being an over-trader increases, but a peak is reached at age 48, beyond which, the likelihood falls again in later years.

Table 4.7 - The effect of background characteristics on the likelihood of being an over-trader

	Income Loss	Income Gain
Variable	**(n=213) (Pseudo R^2=0.102)**	**(n=213) (Pseudo R^2=0.047)**
Income>999euros per month	0.080	-0.024
Gender (Male=1, Female=0)	-0.034	-0.059
Age	**0.048*****	0.022
Age^2	**-0.001*****	-0.000
Married=1, Other=0	-0.07	-0.071
Educated Beyond Minimum School Leaving Age	-0.012	0.042
Have a Degree	**-0.079***	-0.051
Working=1, Not Working=0	-0.082	0.003
Have Insurance	-0.043	-0.016
Have Children	-0.053	-0.052
Religious=1, Atheist=0	0.021	-0.019
VAS Own Health	0.021	0.559
VAS Own Health^2	-0.108	-0.531
40% loss/gain vs 20% loss/gain	0.030	-0.036
60% loss/gain vs 20% loss/gain	-0.007	-0.059
Akaike Information Criterion	0.925	1.058
Hosmer-Lemeshow's Test	0.104	0.259

Values presented are marginal effects coefficients.
Significance is shown as follows: * 10%, **5%, ***1%

4.3.3 Validity

4.3.3.1 Sensitivity to Scale

4.3.3.1.i Sensitivity of the number of years traded to the income change level

Table 4.8 shows the mean number of years respondents were willing to trade, in both the Income Gain and Income Loss questions, *with* and *without* LG respondents. T-tests were performed on the larger sample with LG respondents. These tests were paired when the income change level was the same (e.g. 20% loss vs 20% gain), but unpaired otherwise. T-tests were also performed for the smaller samples but no points of comparison were significant due to the small sample sizes so the results are not presented in the table. Given the survey design sensitivity to scale for each of the two questions can only be tested between individuals, not within individuals.

Looking at the values for the larger sample and looking across the different income change levels, in the Income Loss questions the number of years traded increases as the income change increases. There is no clear pattern in the Income Gain questions. Comparison between 20% and 40% Income Loss proved significant at the 5% level, while comparison between 20% and 60% Income Loss proved significant at the 1% level. There were no significant differences when comparing the Income Gain questions. These results suggest the Income Gain questions were less sensitive to scale than the Income Loss questions.

Considering the smaller sample, excluding LG respondents, in the Income Loss questions the number of years traded increases as the income change level increases. No clear pattern emerges for the Income Gain questions.

4.3.3.1.ii Sensitivity of the number of years traded to the direction of income change

Given diminishing marginal utility of income economic theory would suggest that the Income Loss questions should elicit larger trades than the Income Gain questions. Considering the larger sample in table 4.8, including LG respondents, this is the case for two of the three income change levels. However, these differences are only significant for the

60% income change level (at 1% level). The same relationship holds when considering the smaller sample, but the differences are smaller.

Considering the large sample in Table 4.8, despite the inclusion of LG respondents the standard deviations are high (approximately twice the mean in most cases) which shows the variation in the data. For the Income Loss questions the standard deviations generally increase as the level of loss increases, while no clear relationship can be observed for the gain questions. The median values are 0 in all but one case, which is a product of the large numbers of LG respondents. The 10^{th} and 90^{th} percentiles show the skewness caused by the LG respondents. For the smaller sample, *without* LG responders, the mean, SD and median all increase in all questions.

Table A4.5 in the Appendix shows the mean number of years traded excluding LG respondents and with a balanced panel (i.e. the same respondents occur in the Income Loss and Income Gain questions for each income level). The results show the mean number of years traded to be very similar for the gain and loss questions. There is also strong consistency for the 10^{th} and 90^{th} percentiles. The values are the same for the *gain* and *loss* questions for all three income change levels.

4.3.3.1.iii Sensitivity of the MVQ estimates to the income change level

Sensitivity to Scale can also be assessed by considering the final MVQ estimates generated by the trade-offs made. The MVQ estimates are shown in Tables 4.9 and 4.10. In these tables results are presented based on both individual income and household income. The questions specified 'your current annual income' so it seems probable that respondents will have considered their own individual income. However, some respondents may have thought about their household income, and their decision to trade years may have been a function of household, rather than individual income. Household income was calculated by simply adding an individual's income with their reported partner's income. Table 4.9 shows MVQ values calculated at the individual level, excluding LG respondents. Over-traders are included by truncating negative values to zero. Table 4.10 presents results using the aggregate calculation method, which allows for the inclusion of LG respondents.

Table 4.8 - Mean number of years traded both including and excluding LG respondents

		Variant A: 20% (n=78)		Variant B: 40% (n=80)		Variant C: 60% (n=83)	
		Loss	Gain	Loss	Gain	Loss	Gain
Number of years traded to either avoid an income loss or achieve an income gain *including* LG respondents	Mean	**0.99**	**1.47**	**1.81**	**1.33**	**2.45**	**1.51**
	SD	2.23	2.96	2.74	2.63	3.28	2.89
	10th Percentile	0.00	0.00	0.00	0.00	0.00	0.00
	Median	0.00	0.00	0.00	0.00	1.00	0.00
	90th Percentile	4.92	5.00	5.00	5.50	8.92	5.00
	Paired t-test Loss vs Gain (p-value)	0.143		0.150		**0.009**	
	Unpaired t-test	20% Loss vs 40% Loss	20% Gain vs 40% Gain	40% Loss vs 60% Loss	40% Gain vs 60% Gain	20% Loss vs 60% Loss	20% Gain vs 60% Gain
	p-value	**0.041**	0.763	0.174	0.689	**0.001**	0.933
		Loss (n=22)	**Gain (n=29)**	**Loss (n=37)**	**Gain (n=28)**	**Loss (n=46)**	**Gain (n=30)**
Number of years traded to either avoid an income loss or achieve an income gain *excluding* LG respondents	Mean	**3.5**	**3.95**	**3.91**	**3.81**	**4.43**	**4.17**
	SD	2.99	3.74	2.83	3.24	3.27	3.47
	10th Percentile	0.25	0.25	1.00	0.25	1.00	0.75
	Median	2.25	2.50	4.00	2.96	3.75	2.75
	90th Percentile	9.00	10.00	9.00	10.00	9.92	9.96

T-tests performed on the smaller sample (excluding non-traders) showed no significant differences so are not included in the table.

In Table 4.9 in the Income Loss questions as the income change level increases the MVQ result increases. This is a result of the number of years traded increasing at a decreasing rate. Significant differences were only found when comparing 20% loss with 60% loss (10% level). This is a positive finding and suggests there is a degree of consistency in the MVQ estimates across the income change levels. There is no clear trend in the Income Gain questions. Considering the aggregate results, the MVQ estimates increase as the level of income change increases in both the Income Gain and Income Loss questions. The aggregate results do not allow for statistical testing of differences.

4.3.3.1.iv Sensitivity of the MVQ estimates to the direction of income change

In Table 4.9 the income gain questions produce higher MVQ estimates for two of the three income change levels but these differences are not significant. This is consistent with fewer years being traded in the Income Gain questions. The trend is perhaps reversed for the highest income change level because of the higher income amongst the 60% loss group. The use of a balanced panel (see Appendix A4.6) makes it possible to determine whether differences between the gain and loss results observed in Table 4.9 were the result of different samples being included for the estimates, or a result of differing responses to the two question types. The balanced panel shows that the values generated through the Income Gain questions are now much closer to the values generated through the Income Loss questions, demonstrating that the differences observed in Table 4.9 were a result of different respondents in the two groups, rather than respondents' reactions to the questions. The results of paired t-tests showed no significant differences between values generated through the gain and loss questions.

Considering the aggregate results in Table 4.10, the Income Gain questions produce higher estimates for two of the income change levels.

Table 4.9 – MVQ values calculated at the individual level (excluding LG respondents) based on both individual income and household income

		Variant A: 20%		Variant B: 40%		Variant C: 60%	
		Loss	**Gain**	**Loss**	**Gain**	**Loss**	**Gain**
		(n=22)	(n=29)	(n=37)	(n=28)	(n=46)	(n=30)
Mean number of years traded		3.50	3.95	3.91	3.81	4.43	4.17
Mean Annual Individual Income (Euros)		15,042	16,375	14,834	15,675	21,041	18,630
Mean Annual Household Income (Euros)		27,495	27,409	24,968	25,782	28,910	28,830
Number of negative responses (truncated to zero)		11	17	16	14	13	14
Income equivalent to 1 Quality Adjusted Life Year (Based on Individual Income, Euros)	Mean	**17,439**	**42,212**	**43,564**	**65,957**	**56,827**	**48,846**
	SD	44,561	166,650	138,097	193,760	126,109	108,570
	10th percentile	0	0	0	0	0	0
	Median	0	0	0	1,020	8,673	10,994
	90th percentile	55,944	54,346	65,571	116,683	136,500	98,108
	Unpaired t-test Loss vs Gain (p-value)	0.450		0.606		0.770	
	Unpaired t-test	20% Loss vs 40% Loss	20% Gain vs 40% Gain	40% Loss vs 60% Loss	40% Gain vs 60% Gain	20% Loss vs 60% Loss	20% Gain vs 60% Gain
	p-value	0.294	0.623	0.653	0.683	**0.064**	0.858
Income equivalent to 1 Quality Adjusted Life Year (Based on Household Income, Euros)	Mean	**42,270**	**64,176**	**78,194**	**120,203**	**151,089**	**127,025**
	SD	125,502	216,804	244,851	309,024	450,689	447,923
	10th percentile	0	0	0	0	0	0
	Median	0	0	0	2,398	14,500	12,593
	90th percentile	55,944	64,378	102,000	335,684	273,000	225,300

Table 4.10 - MVQ values calculated at the aggregate level

		Variant A: 20%		Variant B: 40%		Variant C: 60%	
		Loss	**Gain**	**Loss**	**Gain**	**Loss**	**Gain**
Including LG respondents and Over-traders	Number of respondents	78	78	80	80	83	83
	Mean number of years traded	0.99	1.47	1.81	1.33	2.45	1.51
	Mean Annual Individual Income (Euros)	17,471	17,471	15,771	15,771	20,829	20,829
	Mean Annual Household Income (Euros)	27,970	27,970	26,570	26,570	29,587	29,587
	Value of a QALY (Individual Income)	**17,824**	**2,805**	**19,082**	**25,353**	**30,181**	**49,437**
	Value of a QALY (Household Income)	**28,536**	**4,490**	**32,149**	**42,712**	**42,871**	**70,225**

4.3.3.2 Correlation of MVQ estimates with ability to pay

Tables 4.11a and 4.11b show mean QALY values by individual income level for the Income Loss and Income Gain questions respectively. The results are shown for both the individual (excluding LG respondents and including over-traders) and aggregate (including both LG respondents and over-traders) approaches. Looking at the Income Loss values based on the individual approach, there is no clear pattern across the income levels. The highest income level actually produces the lowest MVQ results. This suggests that wealthier respondents value income more highly and hence trade more years to avoid an income loss. For the second highest income level in the 20% Income Loss questions, the mean QALY value is zero. There are only two people in this category and they were both over-traders (and hence their values were truncated to zero). Looking at the aggregate approach for the Income Loss questions, the MVQ results seem to increase as the level of respondent income increases but this relationship is not perfect. Table 4.11b shows that there is no clear pattern in the Income Gain results across different respondent income levels, for either the individual approach or the aggregate approach. However, it should be noted that numbers are very small, especially in the individual approach. It appears that both the Income Loss and Income Gain questions fail this test of validity.

Table 4.11a - MVQ values by individual income level for the Income Loss questions

	Individual Income Level		**20%**	**n**	**40%**	**n**	**60%**	**n**	**Weighted Mean QALY Value**
Individual Approach[1]	Less than 12,000 euros	Mean Years Traded	0.69		0.57		0.65		
		MVQ	**22,523**	11	**13,235**	18	**117,114**	13	**47,821**
	12,000 to 17,999 euros	Mean Years Traded	0.72		0.64		0.52		
		MVQ	**22,800**	5	**104,286**	7	**17,864**	11	**45,239**
	18,000 to 23,999 euros	Mean Years Traded	0.45		0.69		0.58		
		MVQ	**0**	2	**82,579**	7	**56,406**	9	**60,317**
	>24,000 euros	Mean Years Traded	0.55		0.58		0.48		
		MVQ	**5,475**	4	**13,114**	5	**29,801**	13	**21,586**
				22		**37**		**46**	
Aggregate Approach[2]	Less than 12,000 euros	Mean Years Traded	1.18		2.07		1.73		
		MVQ	**5,577**	29	**7,422**	37	**19,662**	26	**10,299**
	12,000 to 17,999 euros	Mean Years Traded	1.00		1.56		2.42		
		MVQ	**15,000**	14	**23,536**	16	**22,190**	22	**20,668**
	18,000 to 23,999 euros	Mean Years Traded	0.69		1.36		2.89		
		MVQ	**40,047**	16	**40,810**	16	**22,674**	13	**35,299**
	>24,000 euros	Mean Years Traded	0.94		1.92		3.08		
		MVQ	**34,748**	19	**33,449**	11	**29,390**	22	**32,206**
				78		**80**		**83**	

1.Excluding LG respondents and including over-traders (truncated to zero)
2. Including LG respondents and over-traders

Table 4.11b - MVQ values by individual income level for the Income Gain questions

	Individual Income Level		20%	n	40%	n	60%	n	Weighted Mean QALY value
Individual Approach[1]	Less than 12,000 euros	Mean Years Traded	0.64		0.62		0.56		
		MVQ	**19,942**	14	**48,449**	11	**68,607**	11	**43,522**
	12,000 to 17,999 euros	Mean Years Traded	0.75		0.72		0.64		
		MVQ	**11,294**	4	**36,571**	7	**36,000**	4	**29,678**
	18,000 to 23,999 euros	Mean Years Traded	0.42		0.59		0.69		
		MVQ	**0**	5	**149,400**	7	**36,120**	5	**72,141**
	>24,000 euros	Mean Years Traded	0.58		0.46		0.53		
		MVQ	**149,965**	6	**4,015**	3	**38,610**	10	**68,312**
				29		**28**		**30**	
Aggregate Approach[2]	Less than 12,000 euros	Mean Years Traded	1.73		1.14		1.86		
		MVQ	**0**	29	**16,977**	37	**13,007**	26	**10,504**
	12,000 to 17,999 euros	Mean Years Traded	0.71		1.22		0.66		
		MVQ	**24,373**	14	**28,221**	16	**112,571**	22	**62,871**
	18,000 to 23,999 euros	Mean Years Traded	1.81		1.81		1.20		
		MVQ	**0**	16	**16,932**	16	**71,488**	13	**26,672**
	>24,000 euros	Mean Years Traded	1.34		1.47		2.12		
		MVQ	**9,000**	19	**40,954**	11	**38,094**	22	**28,069**
				78		**80**		**83**	

1.Excluding LG respondents and including over-traders (truncated to zero)
2. Including LG respondents and over-traders
A shaded box means the value has been artificially truncated to zero (i.e.the actual value is less than zero)

4.3.4 MVQ Results

As stated earlier the aim of this chapter was not to produce definitive MVQ estimates for the Netherlands. However, it is interesting to consider the early signs of the estimates that the methods are likely to produce. Looking at Table 4.9 the mean MVQ result, based on individual income ranges from €17,439 to €65,957. Income levels do vary across the six estimates, with income being noticeably higher in the 60% loss group. Unsurprisingly this group produce the second highest MVQ result. The standard deviations around the results are very high, typically between two and three times greater than the means. This is because as the number of years traded becomes very small the size of the MVQ result increases rapidly (the smallest unit of trade was one month). A respondent with an annual income of €10,000 who traded one month in the 20% Income Loss question, would generate an MVQ value of €230,000. If their income was €30,000 this value would become €690,000. The 10th percentiles and medians are either zero, or very low, owing to the number of over-traders whose values were truncated to zero. The 90th percentile values are generally very high owing to the effect of very small trades. Using household income has the effect of increasing the estimates by the ratio of household income to individual income. However, this effect occurs at the individual level so the relationship in the table is not perfect.

Excluding over-traders (see Appendix A4.7) obviously increases the estimates as these values were truncated to zero. These results range from €34,877 to €131,913. The aggregate results in Table 4.10 are typically much lower than the individual results, as the influence of very small trades is eliminated, ranging from €2,805 to €49,437.

4.3.5 Summary of Results

Two key problems were the existence of LG responses and over-traders. Of all Income Loss responses 56% were LG responses, and this is after excluding the 'extreme non-traders'. Of all the Income Gain questions 64% were LG responses. This poses questions over the feasibility of the method. More employed people, educated people and healthy people were LG respondents in the Income Loss questions. No significant characteristics were found for the Income Gain questions. Few significant variables were found in regression analyses of the determinants of being an LG respondent. For the Income Loss questions employment had a highly significant positive influence in Variant B and in the pooled model. Also a U-shaped age relationship was found in the pooled model, which is consistent with the n-shaped relationship found in the model on the likelihood of being an over-trader. In the Income Gain questions employment also had a significant positive influence for Variant B. The size of the income change level had a negative effect on the likelihood of being an LG respondent in the Income Loss questions, but no effect in the Income Gain questions. Seventeen percent of all Income Loss responses and 19% of all Income Gain responses were over-trades. Age was found to have a highly significant n-shaped effect on the likelihood of being an over-trader in the Income Loss questions. The above significant relationships suggest LG responses and over-trades may on the whole reflect meaningful responses, but this cannot be concluded for individual responses. Regardless of whether the responses are meaningful or protests (or driven by a lack of understanding) it is clear that they are problematic for the calculation of MVQ estimates.

Referring back to the null hypotheses, the first null hypothesis was that the number of years traded in the Income Loss questions would be the same regardless of the level of income loss. The results show that respondents faced with a larger Income Loss on average traded a larger number of years, and these differences were statistically significant on two of three occasions. Therefore, this null hypothesis can tentatively be rejected.

The second null hypothesis was that the number of years traded in the Income Gain questions would be the same regardless of the level of income gain. There was no clear pattern in the number of years traded in the Income Gain questions and no statistically significant differences so this hypothesis cannot be rejected.

The third null hypothesis was that the number of years traded would not differ between the Income Loss and Income Gain questions. There were differences in the number of years traded between the two questions but these were not always in the same direction, and were only statistically significant on one of three occasions, so this null hypothesis cannot be rejected. It was expected that the Income Loss questions would produce larger trades, and this was the case for two of the three income change levels.

Considering the MVQ estimates, the Income Gain questions typically produce largest estimates than the Income Loss questions, which is consistent with smaller number of years being traded. The estimates generated by the Income Loss questions typically increased as the level of income loss increases, but there was no clear pattern for the Income Gain questions.

The fourth null hypothesis was that the MVQ results would not differ depending on respondent income. Breaking MVQ estimates down by respondent income bracket showed no clear pattern for the individual approach. For the aggregate approach the results seemed to increase as the level of respondent income increased, both for the Income Gain and Income Loss questions, but this relationship is imperfect. Therefore we cannot confidently reject the fourth null hypothesis.

4.4 Discussion

The aim of this study was not to present a definitive MVQ for the Netherlands, but to test the feasibility of an alternative method of eliciting an MVQ. The results from this small-scale online study suggest that the Income Gain and Income Loss TTO exercises must overcome a number of problems before they can be considered a viable method for estimating an MVQ. Most notable is the problem of LG responses and over-trades. The individual approach is not able to include the potentially meaningful responses of LG responders. The aggregate approach can do so but this represents a movement away from the welfare economic principle of societal welfare being the sum of individual welfare. The weaknesses of the two questions need to be viewed in light of the fact that the online survey collected poor data across all TTO questions – 25% of respondents did not trade in any of the TTO exercises. An improved study design using face to face interviews may be better able to determine the feasibility of this method.

A serious problem with the TTO based approach is the elicitation of negative MVQ values. Referring to Equation 1, given the assumption of additive separability, a rational respondent should not trade more than two years (i.e. a value of 8 on the right hand side of the equation) because to do so would mean a lower total lifetime income. However, in reality it is plausible that individuals may wish to live for a shorter period of time with high income than for a longer period of time with lower income, even though their total lifetime income may be lower. It is also likely that respondents may not have been able to determine the point at which their lifetime income became lower. If these questions were tested through an interview elicitation procedure it may be possible to use a visual aid that would attempt to make it clearer to respondents in which of the two scenarios lifetime income was higher. This could be done by adapting the standard MVH TTO board (Gudex, 1994) to include an additional strip for lifetime income. This may improve the feasibility of the method.

It would seem that feasibility must be established before a strong assessment of validity can be made. For example, if all LG behaviour represents protest responses then validity is of little importance. However, despite the issue of LG responses the Income Loss questions were still statistically sensitive to scale. This was not the case for the Income Gain questions. Despite the importance of income in the calculation method the MVQ results were not clearly correlated with income.

As touched upon in section 4.2.1.1 the values derived by the method presented in this chapter could be interpreted as the Monetary Value of a Life Year rather than an MVQ since it does not entail a quality of life trade off. The implications of this are that the results should perhaps only apply to life extending treatments. In the broader context of MVQ research, given that previous research has typically produced a range of estimates, and life extensions have been valued more highly than quality of life improvements (E.g. EuroVaQ, Donaldson *et al.* 2011), results from the TTO based method (if proved meaningful) would contribute towards understanding of the upper limit of the MVQ range.

In this study respondents were told to imagine perfect health in both scenarios. In future work it may be preferable to tell respondents they would be in their own current state of health. Their current health could then be valued through either conventional TTO or VAS and the values obtained could be divided by the value of their current health to give MVQ values. This may reduce the number of hypothetical aspects and hence make the task more manageable for respondents who are currently not in full health. However, this approach would entail further dependence upon the assumption of no interactions between health and income. The MVQ value elicited is essentially determined by the choice of income change level. A large scale study would make it possible to gain values for enough income change levels to estimate an indifference curve between health and income and model interactions between the two. An average MVQ value across a range of income change levels could then be estimated.

The results do not suggest a strong case for one type of question over the other (Income Loss vs Income Gain). Comparison of MVQ results generated through the two questions, using a balanced panel, found no significant differences.

4.5 Conclusion

An alternative method for the elicitation of an MVQ based on the TTO has been developed and undergone preliminary testing. A number of problems were encountered, most notably the elicitation of LG responses and negative values. An interview based study that requires respondents to engage in an iterative process, and that can be supplemented by a visual aid may improve the validity of this approach. The next chapter outlines a small scale pilot study that informs the design of a UK based interview study.

CHAPTER FIVE

Further testing of the TTO method of deriving the Monetary Value of a QALY: A pilot study to inform a UK based interview study

5.1 Introduction

The main problems highlighted by preliminary testing of the TTO method of MVQ estimation, as outlined in the previous chapter, were LG responses and over-traders. Although the evidence tentatively suggested these may be meaningful responses, the high number of LG responses across all of the TTO questions suggested the possibility that this may have been driven by the online nature of the exercise. Therefore, to further test the questions in an interview setting, this chapter outlines a small pilot study that preceded the main study, both of which were conducted in the UK. The aim of this pilot was to test and refine both the questions and the visual aid.

In addition to using an interview method of administration with a visual aid, during the course of the pilot study two new questions are included. Rather than asking respondents to trade years of life to either avoid an income loss or achieve an income gain, these new questions ask people to either accept an increase in income to compensate for a fall in life expectancy (TTO3), or give up income to achieve an increase in life expectancy (TTO4). These questions are, therefore, more closely linked to the notions of WTP and WTA. While in the two questions used in the previous chapter years of life were the currency of trade, in the two new questions income is the currency of trade. These two new questions will be examined against the existing WTP literature in the discussion. The next section outlines the methods of the pilot study, before results are presented and implications for the main study drawn.

5.2 Methods

The aim was to test the acceptability of the questions to respondents, in terms of concept, wording and number of questions (burden). In order to achieve this a small number of qualitative interviews were conducted to gain an insight into respondent's thought processes during the exercises. As feedback was gathered design was adapted throughout the pilot study. There were essentially two versions of the pilot. Initially only the Income Loss and Income Gain questions were included and an MVH TTO visual aid was used (Version 1). After six interviews, it seemed that respondents frequently displayed an infinite preference for length of life over income, and two new TTO questions were introduced to further test this (Version 2). Version 2 was also supplemented by a new visual aid.

Ethical consent for both the pilot study and the main study was obtained from the School of Health and Related Research (ScHARR) ethics committee.

5.2.1 Sample Recruitment

Interviews were conducted by the PhD candidate, Carl Tilling. The subjects were admin and clerical staff in Health economics and Decision Science (HEDS) in ScHARR (n=7) and the Department of Economics (n=2), and PhD students in the Department of Economics (n=4). An email was sent out to the ScHARR and Department of Economics admin mailing lists requesting participants. Further emails were sent at later dates to recruit further participants. The PhD students were also recruited through email. No Health Economics PhD students were included in the sample due to concerns they may have insight into this type of exercise that might influence their responses. A total of 13 interviews were conducted. After 13 interviews it was felt that saturation point had been reached i.e. there was little to be gained from conducting further interviews. No incentives were offered for participation.

5.2.2 Preliminary Exercises

Before the interview began respondents were asked to read the information sheet and sign the consent form (see Appendix A5.1 and A5.2).
Prior to the main TTO exercises respondents were asked to complete a number of preliminary exercises, including:

(a) Self-Reported Health on the EQ-5D

(b) VAS of Own Health

(c) TTO of Own Health

(d) Ranking of 7 combinations of years of life and income (listed below)

(e) VAS of the same 7 combinations of years of life and income

Initially exercise (b) asked respondents to place own current health on a scale ranging from best imaginable health to worst imaginable health. However, without dead being valued this information could not be used to rate own current health on a scale from dead to full health. Therefore, after the first 10 pilot interviews the procedure was changed, and the bottom end of the scale was represented by dead. While this rules out the possibility of own current health being valued as worse than dead, it is highly unlikely a member of the convenience sample will consider themselves to be in a state worse than dead.

Exercise (c) used the standard MVH TTO protocol (Dolan, 1997) with a 10 year time horizon. This was supplemented by the MVH TTO visual aid and its protocol to value states worse than dead was also included. However, following the first seven pilot interviews, a modified visual aid was adopted for the main TTO exercises (see later). It was then not practical to use two different visual aids, and require participants to become familiar with both. Therefore, TTO of own health was dropped after the first seven interviews. As there was still an indication of own health through both VAS and EQ-5D it was not felt much would be lost due to this omission.

Exercise (d) asked respondents to rank the following 7 combinations of years of life and income:

• **10 years of life**, in your current state of health, with **100% of current annual income** for each year

• **10 years of life**, in your current state of health, with **80% of your current annual income** for each year

• **10 years of life**, in your current state of health, with **60% of your current annual income** for each year

• **8 years of life**, in your current state of health, with **120% of your current annual income** for each year

• **8 years of life**, in your current state of health, with **100% of your current annual income** for each year

• **7 years of life**, in your current state of health, with **140% of your current annual income** for each year

• **6 years of life**, in your current state of health, with **100% of your current annual income** for each year

These seven combinations were presented to respondents on cards in a random order. Respondents were then asked to place them in order from best to worst, the best being at the top and the worst being at the bottom. These combinations were chosen to reflect scenarios respondents were likely to face in the main TTO exercise, and were included to familiarise respondents with the trade off between years of life and income.

In exercise (e) respondents were asked to place the same seven combinations of years of life and income on a VAS. Initially the top end of the scale was best imaginable life and the bottom end of the scale was worst imaginable life. However, respondents found this too vague and some pointed out that best imaginable life offered no constraint, and they could imagine any number of scenarios that were superior to their first ranked card. Furthermore, valuations were too subjective to attach meaningful values to the cards (although in reality this exercise was included to familiarise respondents with the trade off between length of life and income prior to the TTO exercises). Therefore, after the first three interviews the scale was amended so that the top end was represented by the first ranked card, and the bottom end was represented by 0 years and 0% of income. The card 10 years with 100% of annual income represented 1 (normalisation point), and the other cards could be scaled in relation to this card. The cards were placed either side of the VAS in the order in which they were ranked i.e. 1st ranked would be top left, 2nd ranked would be top right, 3rd ranked would be next left etc. This method was used to avoid the lines from the boxes to the scale crossing too frequently, which can cause confusion for the respondent and difficulty in interpreting the results.

5.2.3 The modified TTO Exercises

Initially, in what will be called version 1 of the pilot study, only two TTO exercises were included, the Income Loss and Income Gain questions included in the Dutch online survey (presented in Chapter 4). The wording of the questions was the same except that rather than being told to imagine being in perfect health, respondents were told to assume they were in their current state of health for the duration of the scenario. As mentioned in the previous chapter, this was done to reduce the number of hypothetical elements in the exercise. Also, although not introducing a direct quality of life trade off into the exercise, it does require respondents to consider quality of life in their valuation which is more aligned with the QALY concept. Therefore, the wording looked as follows:

TTO1: *Trading years to avoid an income loss in current state of health **(Income Loss)***
"You can live for 10 years in your current state of health with (100 - Y%) of your current annual income for each year and then die, or you can live for a shorter period of time in your current state of health with your current annual income for each year and then die."

TTO2: *Trading years to achieve an income gain in current state of health **(Income Gain)***
"You can live for 10 years in your current state of health with your current annual income for each year and then die, or you can live for a shorter period of time in your current state of health with (100 + Y%) of your current annual income for each year and then die."

The "shorter period of time" is varied in these questions until a point of indifference is reached. An MVH TTO visual aid was used with these questions. Coloured cards displayed the amount of income respondents would have in each life, as a percentage of their current income. Consideration was given to creating a visual aid that would display lifetime income in each of the two lives. If over-trading is driven by an inability to comprehend when lifetime income becomes lower a visual aid of this nature may reduce such responses. However, in the Income Gain questions lifetime income is not obvious e.g. if a respondent is faced with 120% income and they trade 1 year, lifetime income becomes 96%. The interviewers in the main study would not be expected to make these calculations. If the visual aid was not going to be used for the Income Gain question it was not feasible to use it for the Income Loss question either.

In version 1 of the pilot study respondents had a total of six TTO exercises to complete. For the first two interviews three income *loss* levels and three income *gain* levels (Y) were chosen at random from the following five values:

20%

40%

50%

60%

80%

Respondents did not necessarily face the same three income change levels in the *gain* and *loss* questions. After the first two interviews it was decided that this method meant respondents could face income change levels that were too similar. Therefore, the procedure was changed so that respondents faced the income change levels 20%, 50% and 80% for both the Income Loss and Income Gain questions. This meant respondents faced a wide range of values and faced both of the extreme values which are most likely to produce LG responses or over-trades. LG responses are most likely to occur when the income gain/loss is smallest, as respondents are unwilling to give up years of their life for small increase in their income, or to avoid a small decrease. However, as the size of the income loss/gain increases respondents may over-trade (see 4.2.3.2 in the previous chapter). In the Income Loss questions respondents would have to give up eight years of life to over-trade in the 80% loss questions, while in the Income Gain respondents would only have to give up 5.56 years to over-trade in the 80% gain questions.

After seven interviews had been completed the frequency at which respondents were not prepared to give up any time (which means an MVQ value cannot be calculated) became a concern. At this stage two additional TTO exercises were introduced, forming what will be called version 2 of the pilot study. These questions asked respondents to either give up income to increase their life expectancy and accept an increase in income to compensate for a fall in life expectancy. The wording of the questions looked as follows:

TTO 3: *Accepting an income gain to compensate for a fall in life expectancy in current state of health **(Years Loss)***

"You can live for 10 years in your current state of health with your current annual income for each year and then die, or you can live for (10-X) years in your current state of health with an income greater than your current annual income and then die".

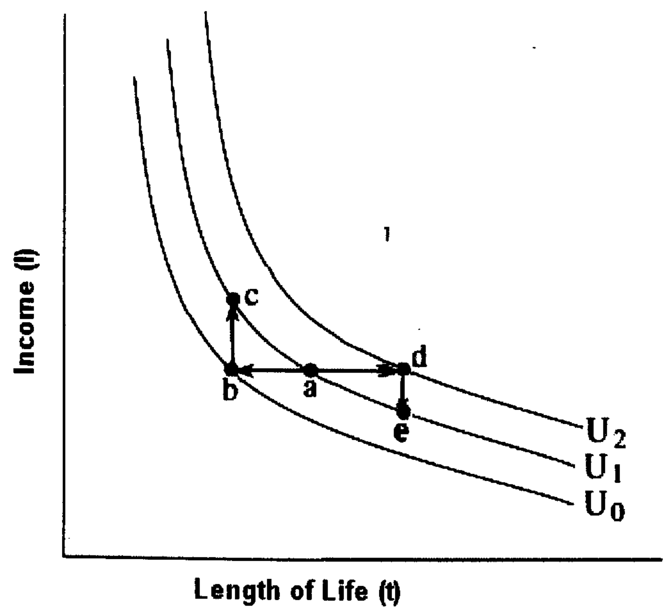
TTO 4: *Trading income to achieve an increase in life expectancy in current state of health **(Years Gain)***

"You can live for 10 years in your current state of health with your current annual income for each year and then die, or you can live for (10+X) years in your current state of health with an income lower than your current annual income and then die".

In TTO3 the "income greater than your current annual income" is varied until a point of indifference is reached. In TTO4 the "income lower than your current annual income" is varied until a point of indifference is reached.

Figure 5.1 represents TTO3 and TTO4 in the context of the theoretical framework set out by Buckingham and Devlin (2006). The x axis represents length of life and the y axis represents income. Each indifference curve represents a level of utility can be achieved by different combinations of longevity and income, where $U_2 > U_1 > U_0$. The starting point is at point *a* on indifference curve U_1 (10 years with current annual income). In TTO3 the respondent is faced with a loss in years which moves them to point *b* on indifference curve U_0. The respondent must then specify the increase in income that will return them to their initial level of utility at point *c* on indifference curve U_1. In TTO4 the respondent is faced (starting point again at point a) the respondent is faced with an increase in life years which moves them to point d on indifference curve U_2. The respondent must then specify the reduction in income that will return them to their initial level of utility at point e on indifference curve U_1.

Figure 5.1 Indifference curve analysis of the Years Loss and Years Gain questions (Adapted from Buckingham and Devlin, p.1151).



TTO3 will henceforth be referred to as Years Loss, while TTO4 will be referred to as Years Gain. The Years Gain questions involve a movement in the same utility space as in the Income Gain questions, but the movement is in the opposite direction. In the Years Gain questions the movement is from a to d to e, while in the Income Gain questions the movement is from e to d to a. The only difference in the Income Loss questions is that respondents start and finish on a lower indifference curve. The Years Loss questions represent the Compensating Variation for a loss and hence differ from the other three question types as they cover different utility spaces. As pointed out by Buckingham and Devlin, 2006) there is no reason to believe that questions eliciting the Compensating Variation for a Loss will produce the same results as questions eliciting the Compensating Variation for a Gain. Any difference between the questions will be determined by the shape of the underlying indifference curves. We might expect the marginal utility of income to diminish at a greater rate than the marginal utility of life years i.e. the indifference curves in the above diagram would be very steep. In this scenario, in the Years Loss questions a very large increase in income would be required to compensate for a fall in life years. Likewise, in the years gain questions a very large decrease in income would be required to return the individual to his/her initial level of utility following an increase in life years. However, responses to the Years Gain questions are constrained by a budget constraint i.e. a respondent can not lose more than all of their income. This could mean that no point of indifference is reached. Alternatively, respondents may feel that there is a minimum amount that they can survive on in order to gain benefit from the additional years.

The inclusion of the Years Loss questions offers a way to examine whether LG responses in the Income Loss and Income Gain questions are meaningful responses or protest responses. An LG response in the Income Loss or Income Gain questions suggests an infinite preference for length of life over income. An LG response in the Years Loss questions is one in which respondents accept no amount of money to compensate for a fall in life years. One might also expect LG respondents to give up all of their income in the Years Gain questions, but this may not occur if respondents have in mind a minimum amount they can survive on. If LG responses in the Income Loss and Income Gain questions are meaningful statements of preference one would expect these respondents to also give an LG response in the Years Loss questions.

The Years Loss and Years Gain questions rely on the same assumptions as the Income Loss and Income Gain questions as outlined in the previous chapter (4.2.1.1).

A new visual aid was created that could be used for all four TTO exercises. As well as a scale for years in each life, the aid also had a scale for annual income in each life. The years scales went up to 20 years and the income scales went up to 200%. All four scales were movable. This visual aid avoided the need for cards as all the necessary information could be displayed with the increased number of scales. An illustration of the visual aid can be found in Appendix A5.3.

5.2.4 Routing in the TTO exercises

In the Income Loss and Income Gain questions both profiles (described as LIFE A and LIFE B) were initially set at 10 years, and respondents were asked to choose between the two. If respondents stated the two were equal, or that they preferred the profile with a lower income they were asked to confirm this preference. If respondents stated that they preferred the profile with higher income (LIFE A), time in that profile was reduced to 5 years, and respondents were again asked to choose between the two. If LIFE A was preferred the duration in LIFE A was reduced, if LIFE B was preferred the duration in LIFE A was increased. Iterations occurred until the respondents stated the two profiles were equal, or changed their preference when moving from one duration to the next. When a respondent changed their preference when moving from one duration to the next they were asked to confirm if the midpoint represented their point of indifference. If they stated a preference for one profile over the other at the midpoint the year would be further subdivided, giving a response accurate to within three months. The only exception to this was if respondents iterated back to 10 years in LIFE A without reaching a point of indifference. In this case they were allowed to trade in units as small as one week. Therefore, a total of seven iterations were required to give an LG response in the Income Loss and Income Gin questions. This is quite different to the Dutch online survey, in which there was no iterative procedure and respondents simply stated the number of years in LIFE A they considered equivalent to 10 years in LIFE B.

The iterative procedure was very similar for the Years Gain and Years Loss questions. In the Years Loss questions respondents initially chose between 10 Years with 100% of annual income (LIFE A) or less than 10 years (determined by choice of X) with 100% of annual income (LIFE B). If LIFE A was chosen the income in LIFE B was increased to 150%. From this point the amount of income was iterated up or down in 10% increments. If LIFE A was still preferred when income was 200% in LIFE B, the respondents was asked if there was any amount of money that would lead them to choose LIFE B. If they said no amount this was classed as an LG response. Therefore, seven iterations followed by an open ended question were required to give an LG response.

In the Years Gain questions respondents initially chose between 10 Years with 100% of annual income (LIFE A) or more than 10 years (determined by choice of X) with 100% of annual income. If LIFE B was chosen the income in LIFE B was reduced to 50%. From this point the amount of income was iterated up or down in 10% increments. An LG responder was someone who still preferred LIFE B with 0% of income. Therefore, seven iterations were required to be an LG responder. In the context of WTP, the procedures for the Years Gain and Years Loss questions are comparable to an iterative bidding game.

5.2.5 Study design

Respondents were given the Years Gain/*loss* in TTO3 and TTO4 that proportionally corresponded to the income *gain/loss* they faced in TTO1 and TTO2 i.e. 20% income loss corresponds to 2 Years Loss, 50% income gain corresponds to 5 Years Gain. However, if all three values for Y were maintained this would have required a further six TTO exercises, giving a total of 12. This was considered too many for a pilot study and the 50% value for Y was dropped, so that just the extreme values were maintained. Therefore the values for Y were 20% and 80% and the values for X were 2 years and 8 years. This meant a total of eight TTO exercises. The participants in the pilot study received no incentive for participation, so it was felt the interview should be kept fairly short. However, in the main study participants received a £25 incentive for participation, hence a larger number of TTOs were included.

5.2.6 Method of calculating an MVQ

The Income Loss and Income Gain questions can be used to estimate an MVQ using the same method as in the previous chapter. The only difference is that since respondents are now told they will live in their current state of health rather than in perfect health, the product of the calculation must be divided by the value of own health (informed by VAS) to gain an MVQ result. VAS of own health was preferred to TTO of own health for two reasons. Firstly, once the new visual aid was employed for the valuation of the hypothetical health/income scenario then an entirely separate visual aid would have been needed to elicit TTO of own health. Secondly, since the majority of respondents are likely to be in good health, there may have been a large number of non-trades in a TTO exercise of own health.

The Years Loss and Years Gain questions can be used to estimate an MVQ in a similar fashion. Consider an individual, in own current health of 0.9 on the VAS scale, faced with Years Loss of 2 years in a TTO3 question, states that 10 years of life with current annual income of £10,000 is equivalent to 8 years of life with 150% of this income, so £15,000. Using prospective lifetime income values and an additive utility function (as in the analysis of the Income Loss and Income Gain questions, see 4.2.3), this point of indifference gives us the following information:

10U (Current Health) + £100,000 = 8U (Current Health) + £120,000	(1)
10U (Current Health) – 8U (Current Health) = £120,000 - £100,000	(2)
2U (Current Health) = £20,000	(3)
U (Current Health) = £10,000	(4)
U (Perfect Health) = £10,000/0.9	(5)
U (Perfect Health) = **£11,111.11**	(6)

While in the Income Loss and Income Gain questions respondents were over-traders if they traded a larger proportion of years than the proportion of income loss/gain they were faced with, in the Years Loss questions they are under-traders if they accept an increase in income that is not large enough to make total lifetime income in the right hand side of equation (1) larger than that in the left hand side. Given the 2 Years Loss example above, respondents are under-traders if they accept an income less than 125%. If faced with 5 Years Loss they cannot accept an income less than 200%. If they are faced with 8 Years Loss they cannot accept an income less than 500%. Since the scale on the visual aid does not exceed 200%, if respondents trade to the end of the scale they are asked: "As you can see this visual aid only goes up as far as 200% of income. If the scale went beyond this value is there any point at which you would choose Life A?" LG responders are those respondents who state that there is no amount of money that would make them choose Life A. It is possible that limiting the scale to 200% will make respondents more reluctant to go above 200%, and hence increase the likelihood of them being under-traders. However, to increase the scale as far as 500% would either have required making the aid unmanageably large, which would have been unpopular with interviewers, or condensing the scale to an extent that would jeopardise the clarity of the exercise for respondents.

The effect of an increase in the amount of income required to compensate for a fall in life years is complex. Using the example above, and assuming own health is at perfect health (i.e. 100 on the VAS scale), if the respondent required 200% of current income rather than 150% the MVQ result would increase from £10,000 to £30,000 (an increase of £20,000). This is because the ratio of the number of years remaining (8) to the number of years lost (2) is 4:1. The increase in the amount of compensation required is 50%. 50% of annual income is £5,000. Multiplying this by the ratio above gives £20,000, which is the size of the increase in the MVQ result. Consider another example. If the respondent required 250% of current annual income in order to live for only 5 years he/she would generate an MVQ result of £5,000. If the amount of income required increased to 300% the MVQ result would increase to £10,000 because the ratio of years left to years given up is 1:1. The amount of income required has increased by 50% so the MVQ result increases by 50% of annual income (£5,000). So, an increase in the amount of income required will increase the MVQ result by the ratio of years left to years lost multiplied by the amount of the increase in income required. In other words the larger the number of years lost the smaller the impact of an increase in compensation required on the MVQ result.

Respondent income has a proportional effect on results i.e. doubling respondent income will double the MVQ result. Doubling the increase in income required to compensate for a doubling in the number of years lost will halve the size of the MVQ result.

Now imagine the same individual facing Years Gain of 2 years in a TTO4 question states that 10 years of life with a normal annual income of £10,000, is equivalent to 12 years of life with 70% of this income, so £7,000. This point of indifference gives us the following information:

10U (Current Health) + £100,000 = 12U (Current Health) + £84,000 (7)

12U (Current Health) – 10U (Current Health) = £100,000 - £84,000 (8)

2U (Current Health) = £16,000 (9)

U (Current Health) = £8,000 (10)

U (Perfect Health) = £8,000/0.9 (11)

U (Perfect Health) = **£8,888.89** (12)

Respondents in the Years Gain questions are under-traders (i.e. produce a negative MVQ value) if they do not give up enough income to make total lifetime income in the right hand side of equation (7) lower than in the left hand side. With 2 Years Gain as outlined above they must allow their income to fall below 83.3%. If faced with 5 Years Gain they must allow their income to fall below 66.7%, and if faced with 8 Years Gain this figure becomes 55.6%. A difference between TTO4 and the other TTO questions is that the MVQ results are severely constrained by personal income, and the larger the Years Gain presented to respondents the smaller the results will be. In the example above, if the respondent had been in perfect health and given up all their income the MVQ result would only be half of lifetime income based on full annual income (i.e. £50,000). If they were faced with 5 Years Gain and once again gave up all of their income the MVQ result would only be 20% of lifetime income based on full annual income (i.e. £20,000). In order for the definition 'LG responder' to maintain the same meaning (i.e. someone who values life years infinitely more highly than income), LG responders in TTO4 are respondents who give up all of their income. In contrast to the other TTOs these responses are not problematic and can be included in all MVQ calculations.

An increase in the amount of income given up increases the MVQ result by the ratio of the number of life years with the gain (12 in the example above) to the number of Years Gained (2) i.e. 6 in the example above. Therefore if the amount of income required with the higher life expectancy falls from 70% to 60% (example above) the MVQ result will increase from £8,000 (assuming perfect health) to £14,000. In other words, the larger the number of years to be gained the smaller the effect of an increase in the amount of income given up. An increase in respondent income has a proportional effect on the MVQ result. Doubling the amount of income given up in response to a doubling of the number of years to be gained will lead to a higher MVQ result.

5.2.7 Feedback Questions

After all the TTO exercises respondents were asked the following eight open ended feedback questions:

1. What were your main considerations when answering these questions?
2. Do you consider it unethical to give up years of life for an increase in income?
3. [Only asked if respondent did not trade in any of the exercises] Why were you not prepared to trade any time for an increase in income in this question?
4. On a scale from 1 to 5, how clear did you find the wording of the questions, where 1 is not very clear and 5 is very clear?
5. How do you think the wording of these questions could be improved?
6. On a scale from 1 to 5, how useful did you find the visual aid, where 1 is not very useful and 5 is very useful?
7. How do you think the visual aid could be improved?
8. [If negative value given in Income Loss] Did you realise in this question that your lifetime income would be lower?

The questions were asked, and responses noted, by the interviewer. The questions were also audiorecorded. Question 8 was soon dropped (after two interviews) as it was too complicated for respondents to understand. Questions 5 and 7 also proved uninformative as most respondents had very little to contribute. There was a blank page at the end of the respondent booklet for any comments or feedback they had.

5.2.8 Background Characteristics

The background questions asked of respondents at the end of the pilot interviews are outlined in Table 5.1.

Table 5.1 Background characteristics collected at the end of the pilot interviews

Background Question	**Categories**
Gender	Male/Female
Age	Open-ended
Did your **education** continue after the minimum school leaving age?	Yes/No
Do you have a degree or equivalent professional qualification?	Yes/No
Marital Status	Single/ Married/ Co-habiting/ divorced/ widowed
Religion	Roman Catholic, Protestant, Muslim, Hindu, Jewish, Atheist, Other
Main Activity	Employed, Self-Employed, Retired, Housework, Student, Seeking Work, Other
Home Ownership	Own Home outright or with a mortgage, rent from local authority, rent from private sector
Net Own Annual Income	Chosen from 10 income brackets, ranging from "I do not have an income" to >£50,000
Net Partners Annual income	Chosen from 10 income brackets, ranging from 'My partner does not have an income' to >£50,000.
Number of people in the household	
Number of these under the age of 18	

5.2.9 Feedback from the interviewer

At the end of the interview the interviewer (CT) noted whether he agreed or disagreed, on a scale from 1 to 5 (where 1 is strongly agree and 5 is strongly disagree), with each of the following statements:

1. "The respondent seemed to understand the questions well"
2. "The respondent appeared to lose interest when valuing the latter income levels"
3. "The respondent gave a lot of thought to the decisions he/she was asked to make"

There was a blank page at the end of the interview booklet for any comments or feedback the interviewer had.

5.2.10 Life Satisfaction Questions

For the very last interview life satisfaction questions were included. It may be that people's perceptions of aspects of their life, such as health and income, are more important than their actual health and income. For example, someone's income may be relatively low, but may be high in relation to their peers that they compare themselves to. This individual may therefore be unwilling to trade many years to increase his income. The questions were inserted after the feedback questions and before the background questions. The questions were taken from the British Household Panel Survey (available through the Institute for Social and Economic Research, www.iser.essex.ac.uk/bhps). The questions ask how satisfied someone is, on a scale from 1 to 7 (where 1 is not satisfied at all and 7 is completely satisfied), across the following domains:

1) "Your Health"
2) "The income of your household"
3) "Your job (if in employment)"
4) "Your social life"
5) "The amount of leisure time you have"
6) "Your life overall"

5.3 Results

The main interest in the pilot study is the number of LG responders and over-trades/under-trades in the TTO exercises and how the questions were received by respondents, as indicated by the feedback questions. The sample size is too small to calculate a reliable MVQ estimate. Therefore, this section presents background characteristics of the sample, the prevalence of LG responses and over-trades/under-trades in the four TTO questions and the common responses to the feedback questions. The implications for the main study are then discussed.

5.3.1 Background Characteristics

Table 5.2 shows the background characteristics of the pilot sample. The vast majority of the sample (11 out of 13) are female. The average age of the sample, at 37.1, is lower than the national average. This is unsurprising given that PhD students formed part of the sample, and there were no pensioners. The sample is very well educated, which is also a result of recruiting PhD students. The sample consists only of employed people and students. There is a relatively good spread of income levels, although there is no one in the two highest income brackets. It should be noted that current income may be a poor indication of permanent income, or socioeconomic class, for PhD students. It was also observed that respondents had difficulty specifying their net annual income to the level of accuracy required by the ten income brackets, and that most respondents seemed to consider their income on a monthly rather than annual basis. More than half of the sample are atheist. The sample is evenly split between divorced, married and single. Just under half of the sample are homeowners. The majority of respondents reported to be in the best EQ-5D state 11111.

For version 1 (two TTO questions with standard visual aid) of the study the average completion time was 25 minutes. For version 2 (four TTO questions with new visual aid) the average completion time was 29 minutes. The overall average was 26.7 minutes.

Table 5.2 - Background Characteristics of the pilot study sample

Number of Respondents		13
Gender	Male	2
	Female	11
Age	Average	37.1
	SD	13.6
Educated beyond min school leaving age	Yes	11
	No	2
Educated to Degree Level	Yes	6
	No	7
Marital Status	Married	4
	Single	4
	Co-habiting	1
	Divorced	4
	Widowed	0
Religion	Roman Catholic	1
	Protestant	2
	Atheist/Agnostic	7
	Other	3
Main Activity	Employed	10
	Self-Employed	0
	Retired	0
	Housework	0
	Student	3
	Seeking Work	0
	Other	0
Home Ownership[1]	Own/mortgage	6
	Rent	6
Net Own Annual Income	I do not have an income	0
	<£5,000	1
	£5,000-£10,000	0
	£10,000-£15,000	4
	£15,000-£20,000	4
	£20,000-£25,000	2
	£25,000-£30,000	1
	£30,000-£40,000	1
	£40,000-£50,000	0
	>£50,000	0
Net Partners Annual Income	I do not have a partner	8
	My partner does not have an income	1
	<£5,000	0
	£5,000-£10,000	0
	£10,000-£15,000	1
	£15,000-£20,000	0
	£20,000-£25,000	0
	£25,000-£30,000	1
	£30,000-£40,000	1
	£40,000-£50,000	0
	>£50,000	1

1. One subject reported that she lived at home with her parents

Table 5.2 Continued		
Number of People in the Household	1 2 3 4 7	4 3 4 1 1
Children under 18 in the household	Yes No	3 10
EQ-5D Own Health	Number in 11111 One level 2 dimension Two level 2 dimensions	9 2 2
Mean Completion Time		26:41

5.3.2 The TTO Results

Table 5.3 shows the number of LG responses and over-trades for the Income Loss and Income Gain questions, for the full sample and for the two different versions. Version 1 corresponds to the initial procedure including only the two TTO exercises and using a standard MVH TTO board. Version 2 corresponds to the amended procedure including the two new TTO exercises and using the new visual aid. Table 5.3 shows that in the Income Loss and Income Gain questions LG responses are a far greater problem than over-trading. In total 43 out of 70 responses (61%) were LG responses. This perhaps suggests that the large number of LG responses in the online Dutch study were caused by the actual questions rather than by the mode of administration. LG responses were slightly more prevalent in the Income Gain questions than the Income Loss questions. This was also found in the online Dutch survey. Over-trades only occurred on seven occasions. All over-trades occurred in the Income Gain questions. Unfortunately the new visual aid seems to have had little effect on the prevalence of LG responses and over-trades. In fact, there were more over-trades in version 2 of the interview.

Table 5.3 - Number of LG responses and over-trades for the Income Loss and Income Gain questions

	INCOME LOSS					INCOME GAIN				
Level of Income Loss/Gain	20%	50%	80%	40%	60%	20%	50%	80%	40%	60%
Overall (n=13)										
Number of responses	13	6	12	2	2	12	7	12	2	2
Number of LG responses	8	4	6	1	1	8	5	7	1	2
Number of over-trades	0	0	0	0	0	3	1	3	0	0
Version 1 (n=7)										
Number of responses	7	6	6	2	2	6	7	6	2	2
Number of LG responses	4	4	4	1	1	4	5	4	1	2
Number of over-trades	0	0	0	0	0	1	1	1	0	0
Version 2 (n=6)										
Number of responses	6		6			6		6		
Number of non trades	4		2			4		3		
Number of over-trades	0		0			2		2		

Table 5.4 shows that in the Years Gain questions there was only one problematic response. For 2 Years Gain one respondent under-traded i.e. they would not allow their income to fall below 83.3% in order to achieve an increase in life expectancy by two years. In the Years Loss questions nearly all responses are LG responses (10 out of 12) i.e. there was no amount of money that would make respondents choose the shorter life expectancy. There was also one under-trade i.e. for 2 Years Loss a respondent would accept an income lower than 125% to compensate for the fall in life expectancy. Therefore, there was only one unproblematic response in the Years Loss questions. Given the apparent preference for life over money in the Years Loss questions one might expect respondents who have given LG responses in these questions to do likewise in the corresponding Income Gain questions. The numbers in brackets show the number of respondents that have done this. Of the four LG responders in the 2 Years Loss questions, all of them gave LG responses in the corresponding 20% Income Gain questions. Of the six LG responders in the 8 Years Loss questions, three of them gave LG responses in the corresponding 80% Income Gain questions.

Table 5.4 - Number of LG responses and under-trades for the Years Loss and Years Gain questions

	YEARS LOSS		YEARS GAIN	
Number of years lost/gained (X)	2	8	2	8
Number of responses	6	6	6	6
Number of LG responses	4(4)	6(3)	0	0
Number of under-trades	1	0	1	0

Numbers in brackets show respondents that have non-traded in the Years Loss questions and also under-traded in the corresponding Years Gain question

5.3.3 Feedback Questions

Summaries of the common themes in responses to the open ended feedback questions are presented here. The full list of verbatim responses can be found in Appendix A5.4.

1. What were your main considerations when answering these questions?

All except one of the respondents gave responses that were indicative of a strong preference for years of life over income. Three of these indicated that they wanted to be able to spend time with their family. Two respondents indicated that they could rely on their husband's income which made the decision easier for them. Four respondents indicated that although they had a strong preference for life years, they would want to avoid poverty. The one respondent that did not indicate a strong preference for life years simply replied "the money". In summary, spending time with family, whether or not a respondent can rely on their partner's income, and a desire to avoid poverty appear to be the most important motivators behind respondent's trading decisions.

2. Do you consider it unethical to give up years of life for an increase in income?

Most respondents only gave a one word response to this question. Only four out of thirteen respondents thought that giving up years of life for an increase in income was unethical. One of these four respondents acknowledged the contradiction they had made, since they had traded in the Income Loss and Income Gain questions. Of the four respondents who elaborated upon their 'no' responses, three of them argued it was up to people to make their own decisions. In summary, responses to this question suggest the TTO questions dido not elicit as many ethical objections in this sample as had been anticipated.

3. Why were you not prepared to trade any time for an increase in income in some of the questions? (Relevant for 10 of 13 respondents)

Responses to this question further confirmed the findings in question 1: i.e. respondents have a strong preference for years of life over income. Some respondents stressed the importance of years of life, while others stressed the unimportance of money.

General Comments:

Of the six respondents who gave general comments, four of them indicated that the VAS with the different combinations of years of life and income were the hardest part of the exercise. All of these respondents completed were given the VAS with the modified scale (where 0 correspondents to 0 years of life with 0% of income, and 10 years with 100% of income corresponds to 100), suggesting the modified scale has not succeeded in making the task easier for respondents. One respondent indicated it would be easier if a specific monetary amount were used, rather than percentages. One respondent indicated that if you lived one life you wouldn't know that the other life was an option, so you would "get on with what you're dealt".

For the first version of the pilot study, with only the Income Loss and Income Gain questions, the clarity of the questions was scored, on average, as a 4 (where 5 is the best and 1 is the worst). For the second version of the study, including the Years Gain and Years Loss questions, the clarity of the questions was scored, on average, as a 4.5. The overall average was 4.23.

For the first version of the study, the usefulness of the visual aid was scored, on average, at 4.07. For the second version of the study, the usefulness of the aid was scored, on average, at 4.5. The overall average was 4.27.

5.4 Discussion and Conclusion

The aim of this pilot was to develop and improve the protocol for use in the main study. The pilot suggested that respondents can understand the tasks well and complete them in a relatively short period of time. A number of changes were made to the protocol as the pilot study proceeded, including the addition of two new TTO questions and the development of a new visual aid. According to the feedback questions these changes seem to have, if anything, improved the clarity of the exercise. However, some caution is warranted since the sample used in the pilot study was better educated than a representative sample.

The Years Gain and Years Loss questions are essentially WTP and WTA questions respectively. As shown in the literature in Chapter 3, WTA questions typically elicit higher values than WTP questions. The findings from this pilot study, that LG responses are more prevalent in the Years Loss questions, are consistent with this. This finding is likely due to the influence of the budget constraint in Years Gain questions (WTP), an effect that has been observed previously in the WTP literature. Further comparison between the Years Gain and Years Loss questions and the WTP/WTA literature will be made in the next chapter.

The main concern arising from the pilot study was the high prevalence of LG responses in the Income Loss, Income Gain and Years Loss questions. The consistency between the questions suggest that these responses may represent meaningful preferences for length of life over income, but this needs to be tested further in the main study. A large number of respondents in the pilot study were married women, who could rely on their husband's income if necessary, which further increases their preference for years of life over income. Using a more representative sample of the general public in the main study may reduce the prevalence of LG responses.

A number of factors suggest that LG responses in the Income Loss and Income Gain questions, as observed in both this pilot study and the previous Dutch online survey, are meaningful statements of preference rather than protest responses. The majority of respondents did not find the questions unethical, but PhD students may be more tolerant than the typical member of society. The majority also stated a strong preference for life years over income. The responses to the two new questions, Years Loss and Years Gain, also further confirmed the preference for years of life over income.

A recurring theme amongst the feedback from participants was that the VAS of the seven different combinations of years of life and income was the most difficult part of the exercise. However, the interviewer (CT) did observe that this exercise did force the respondents to give serious consideration to the relative importance of life years and income. Therefore, this exercise is retained in the main study. When training the interviewers for the main study it was going to be important to emphasise to them that thorough explanation of this exercise was required. The use of VAS is not ideal as it means own health is not measured on the same scale as the hypothetical values. While VAS of own health is maintained in the next chapter due to a desire to avoid having multiple visual aids, future work should seek to elicit valuations of own health through TTO.

A number of changes have been made throughout the pilot study that will inform the design of the main study in the next chapter. The anchoring of the VAS exercise has been refined, a new visual aid has been developed and tested, life satisfaction questions have been introduced, and two new TTO based questions akin to standard WTP and WTA questions have been tested.

CHAPTER SIX

Further testing of the TTO method for estimating the Monetary Value of a QALY: A UK based interview study

6.1 Introduction

This chapter presents the UK-based main interview study. The aim of the study is to further test the four TTO questions used in the pilot study to see if the problems of LG responses and over-trading, observed in the Dutch online survey and in the pilot study, persist in interviews with members of the general public. The methods will be assessed in terms of their feasibility, reliability and validity to determine whether any of the four TTO questions are a credible alternative to WTP for the derivation of an MVQ, and hence whether they can be recommended for use in further research.

6.2 Methods

6.2.1 Sample Recruitment

Interviews were conducted by a private data collection company based in Oxford, Oxford Outcomes. Three different interviewers were used to recruit a sample of 100 members of the general public. This sample size was dictated by the available budget. While this does not give strong power in statistical tests the main aim of this study was to test the feasibility of the TTO questions in face to face interviews, rather than to present definitive MVQ values. The interviewers had a pool of participants that had participated in previous surveys which they could draw upon. These participants might be asked if they knew anyone who would like to participate to aid further recruitment through a 'snowballing' approach. Given the rather unique nature of the TTO questions in this study it was not felt that previous experience of TTO would pose a problem. Some participants were also approached on the street, and then taken to a quiet location if they agreed to participate. An incentive of £25 was offered for participation.

Interviewers were instructed to recruit only domestic citizens (not tourists), and to aim for respondents of working age. It was felt that people of working age were most relevant for questions regarding income, and it was felt that these people would be more able to imagine a scenario in which their income fell (e.g. as a result of unemployment). They were also instructed to recruit a sample as representative of the population as possible. The interviewers were trained in how to use the visual aid and the interview script in a two hour session administered by CT.

6.2.2 General Design

The outline of the questions in this section is presented in the order in which they appeared in the interview. Initially preliminary exercises included EQ-5D measurement of health and VAS of own health. A short ranking exercise was then performed. This was followed by the TTO exercises. Following these exercises respondents were asked to complete feedback questions, life satisfaction questions, income expectations questions, and background characteristics. Following the interview the interviewers rated the respondent's understanding.

6.2.2.1 Preliminary Exercises

Prior to the start of the interview respondents were asked to read the information sheet and sign the consent form (see Appendices A5.1 and A5.2). During interviewer training some of the interviewers were concerned that respondents would not want to give an indication of their income. Since this information is crucial to the estimation of an MVQ, and since background characteristics appear at the end of the exercise, it was decided that respondents would be told at the start of the interview that they would be required to give an indication of their salary. If they were not prepared to do this the interview was not started.

The preliminary exercises, prior to the main TTO exercises, were predominantly the same as those used in the pilot study. First was self-reported health on the EQ-5D. Second was VAS of own health, results of which are used to inform the MVQ estimates (as in equations 5 and 11 in Chapter 5). The VAS scale was anchored at "best imaginable health" and "worst imaginable health". So that values could be scaled from 0 to 1, where 0 is dead and 1 is full health, respondents were asked to place immediate death on the scale. Unlike in the pilot study, this allows for the possibility of a health state worse than dead. In the convenience sample in the pilot study it was highly unlikely that a respondent would have been in a state worse than dead. While this is also unlikely in the main study, there seemed no reason to rule this possibility out. A negative value for own health is problematic for the estimation of an MVQ since the monetary amount in equations 5 and 11, will be divided by a negative number, hence giving a negative estimate. However, it was not anticipated that any respondents will value their own health as worse than dead.

The third preliminary exercise was ranking of the seven combinations of years of life and income used in the pilot study. The final preliminary exercise was VAS of the same seven cards. The scale extended from 0 to 130 (see Appendix A6.1 for visual presentation of the scale and wording of the instructions). The cards were placed either side of the scale in the order in which they were ranked. The bottom end of the scale was marked as "0 years of life with 0% of income". A line was drawn by the interviewer from the card representing "10 years of life with 100% of annual income each year" to 100 on the scale. Respondents were then asked to value the other six cards relative to these two points. If respondent wished to value a card as higher than 130, the interviewer asked them what value they wanted to give the card and made a note accordingly. This procedure differed slightly from that used in the pilot study. In the pilot study the scale only extended to 100, and the first ranked card in the ranking exercise represented 100. However, it was noticed that this ruled out preference reversals to some degree (i.e. inconsistency between the ranking and VAS exercise) which may bias results. The procedure in the main study does not impose that the first ranked card be valued highest in the VAS exercise. The number 130 as the upper limit of the scale was chosen as it was felt this would give most respondents enough space to state their preferences, but would not jeopardise the clarity of the scale.

6.2.2.2 The TTO exercises and Interview Variants

All four TTO questions used in version 2 of the pilot study were included: Income Loss, Income Gain, Years Loss and Years Gain. The wording of the questions was exactly the same as in the pilot. Three levels of income Gain/Loss (Y) were used: 20%, 50%, 80%. Three levels of Years Gain/Loss *(X)* were also used: 2, 5, 8. Respondents answered all four TTO questions, with all three values of X and Y, giving a total of 12 TTO questions per respondent. The order in which respondents received the different TTO questions was varied, and there were four variants:

Variant A: Income Loss, Income Gain, Years Loss, Years Gain

Variant B: Income Gain, Income Loss, Years Gain, Years Loss

Variant C: Years Loss, Years Gain, Income Loss, Income Gain

Variant D: Years Gain, Years Loss, Income Gain, Income Loss

The order in which respondents received the different levels of X and Y for a particular TTO question was randomised by shuffling three cards. The new visual aid (see Appendix A5.3) used in version 2 of the pilot study was used with all TTO questions. The iterative procedure was as presented in 5.2.6.

6.2.2.3 Feedback Questions

It would not have been practical to use open ended feedback questions when the data collection was contracted out. The accuracy of the interviewers' noting of responses would have been a concern, and there may have been bias introduced by the use of multiple interviewers. Furthermore, open ended questions would have increased the length of interviews considerably which may have jeopardised the sample size, given existing budget constraints. Therefore, respondents were asked to indicate whether they agreed or disagreed with eight statements, by circling a number from 1 to 5, where 1 represents strongly agree and 5 represents strongly disagree. The eight statements were based upon common responses in the pilot study:

1. "The **visual aid** in this survey was very helpful"

2. "If my income fell I could rely on my **partner's income**" (*Only relevant if respondent has a partner with an income*)

3. "If I didn't have **children** I would have answered these questions differently" (*Only relevant if respondent has children*)

4. "When answering these questions I had in mind a minimum amount of money that I could **survive** on"

5. "I consider it **unethical** to give up years of life for an increase in income"

6. "Time is more important than money because I want to be able to spend time with **friends and family**"

7. "When answering these questions I attempted to calculate the total **lifetime income** I would have in the two scenarios"

8. "I found it **difficult to imagine** a scenario in which my income would fall as much as is specified in the questions."

In questions 2 and 3 "NA" was included as an option for those who did not have a partner with an income or did not have children. These questions were included to enable analysis of the relationship between responses and TTO valuations.

At the end of the participant booklet a blank page was included to allow respondents to note any additional comments they had about the questions they had been asked. Respondents were asked to write these comments themselves.

6.2.2.4 Life Satisfaction

The same six life satisfaction questions that were included in the final pilot interview were included in the main study. The questions ask how satisfied someone is, on a scale from 1 to 7 (where 1 is not satisfied at all and 7 is completely satisfied), across the following domains:

1) "Your Health"
2) "The income of your household"
3) "Your job (if in employment)"
4) "Your social life"
5) "The amount of leisure time you have"

6) "Your life overall"

6.2.2.5 Income Evaluation Question

Respondents' satisfaction with their income may be more important than their actual income in explaining their willingness to trade years of life to achieve an increase in income. For this reason, some open ended questions on income expectations were included (Van Praag and Ferrer-i-Carbonell, 2007). Respondents were asked:

"Please try and fill in the amounts asked for to the best of your judgement. Taking into account my present living circumstances I would regard a **net monthly household income** as:

Excellent if it were above £__________

Sufficient if it were between £__________ and £__________

Very bad if it were below £__________"

6.2.2.6 Background characteristics

The background questions were the same as in the pilot. The only difference was that the income brackets were presented as both weekly and monthly, but not annual, as it was felt this was more relevant. A number of respondents in the pilot study seemed to find it difficult to calculate their income in annual terms, and suggested monthly income was more relevant. Also, the majority of employees in the UK are paid on a monthly basis. Only seven income brackets were used (including "I do not have an income") rather than the 10 income brackets used in the pilot. Income was also specified on a weekly and monthly, rather than annual, basis. It was felt this would make it easier for respondents to identify their appropriate category. The income brackets for both net own income and net partner's income were:

Weekly	Monthly
I do not have an income	
<£100	<£400
£100 - £300	£430 - £1300
£300 - £500	£1200 - £2200
£500 - £700	£2000 - £3000
£700 - £1000	£3000 - £4500
>£1,000	>£4,500

6.2.2.7 Feedback from the interviewer

As in the pilot study the interviewer was asked to state whether they agreed with some statements. However, the scale ranged from 1 to 3 (where 1 is agree and 3 is disagree) rather than 1 to 5 as was the case in the pilot. When conducting the pilot interviews CT found it difficult to grade the statements to the degree of accuracy required by a scale ranging from 1 to 5. The three statements that were included in the pilot were included in the main study, as well as two additional statements, giving a total of five:

1. "The respondent seemed to understand the questions well"
2. "The respondent appeared to lose interest when valuing the latter income levels"
3. "The respondent gave a lot of thought to the decisions he/she was asked to make"
4. "The respondent needed a lot of help from me"
5. "The interview was interrupted by others or the respondent was distracted"

6.2.3 Analysis and Null Hypotheses

6.2.3.1 Background Characteristics

Background characteristics are presented for the full sample and compared with UK population values. Background characteristics are also presented by questionnaire variant. There are four variants of the interview with different TTO orderings in each. Chi^2 tests are used to determine whether there are any significant differences between the variants.

6.2.3.2 Feasibility

Firstly, the interview completion rate is presented. The number of LG responses and over/under traders are then presented for each TTO question and for each level of years/income loss/gain. Analysis is then performed to identify any determinants of being an LG responder or over/under trader. For this analysis LG responders are defined as respondents who specify an infinite preference for length of life over income for all three income/years levels in a given TTO. Over/under traders are defined as respondents who over/under trade in at least one of the income/years levels in a given TTO. Therefore, a respondent can be an LG responder in one TTO, and an over/under trader in another TTO. Chi^2 tests are used to compare background characteristics by LG responders and the rest of the sample, and by over/under traders and the rest of the sample, by TTO question. Mean responses to the life satisfaction questions are compared (through t-tests) by LG responders and the rest of the sample, and by over/under-traders and the rest of the sample. Mean responses to the income expectations questions are compared (through t-tests) by LG responders and the rest of the sample, and by over/under traders and the rest of the sample. Chi^2 tests are also used to compare the proportions of respondents whose incomes exceed their stated income expectations values. Finally, mean responses to the feedback questions are compared (through t-tests) by LG responders and the rest of the sample, and by over/under traders and the rest of the sample.

6.2.3.3 Reliability

Internal reliability is assessed through a test of logical consistency. For each question type we would expect respondents' valuations of each of three income/years change levels to give a logical ranking. For example, in the Income Loss questions we would expect respondents to trade more years to avoid a 50% income loss than they would to avoid a 20% income loss, and to trade even more years to avoid an 80% income loss. Similarly, in the Income Gain questions we would expect respondents to trade more years to achieve a 50% Income Gain than they would to achieve a 20% Income Gain, and to trade even more years to achieve an 80% Income Gain.

To assess logical consistency respondents are placed into one of four categories: 'total LG responder', 'consistent', 'weak consistent' and 'weak inconsistent'. 'Total LG responders' are respondents who gave LG responses in each of the three income/years change levels. 'Consistent' respondents are those whose valuations assumed the expected ranking. 'Weak consistent' are those respondents who valued two of the income/years change levels as equal but then valued the third level differently and in the expected direction. Finally, 'inconsistent' respondents are those whose valuations contradicted logical consistency, or who traded the same number of years/income for each of the three income/years change levels. The analysis is performed for the Income Loss, Income Gain and Years Gain questions. It is not performed for the Years Loss questions due to the very high number of LG responses.

6.2.3.4 Validity

As in Chapter 4 validity is assessed through sensitivity to scale and correlation with ability to pay. In addition, a simple test of convergent validity is performed using the results from the ranking exercise.

6.2.3.4.i Sensitivity to Scale

Sensitivity to scale is assessed firstly by considering the responsiveness of the number of years (income) traded to the size of the income (years) change level. This gives rise to the first four null hypotheses:

Null Hypothesis 1: *The mean number of years traded in the Income Loss questions will not differ depending on the income change level*
See section 4.2.3.3.i for explanation.

Null Hypothesis 2: *The mean number of years traded in the Income Gain questions will not differ depending on the income change level*
See section 4.2.3.3.i for explanation.

Null Hypothesis 3: *The mean amount of income traded in the Years Gain questions will not differ depending on the number of years to be gained*
Alternatively, given diminishing marginal utility of income, and the influence of the budget constraint, we would expect respondents to give up an increasing amount of income at a decreasing rate.

Null Hypothesis 4: *The mean amount of income given traded in the Years Loss questions will not differ depending on the number of years lost*
Alternatively, given both diminishing marginal utility of income and life years, we would expect the amount of income given up to increase at an increasing rate.

The mean number of years traded in the four different questions is presented for each of the income/years change levels. The values for the Income Loss and Income Gain questions are presented with LG responses. In the Years Loss questions no value can be attached to

LG responses i.e. they give a qualitative 'no amount' response. In the Years Gain questions LG responses are not problematic (they are defined as respondents who give up all of their income, and can be used to estimate an MVQ). The mean number of years/amount of income traded across the different years/income change levels is compared through paired and unpaired t-tests. This allows the above four hypotheses to be tested formally.

The sensitivity of the MVQ estimates to the income (years) change level is also tested. The nature of the calculation method means we would expect the MVQ estimates to be less sensitive to scale than the number of years traded. For example, in the Income Loss questions if the number of years traded increased proportionately with the level of income loss the MVQ estimate would remain constant. In the Income Gain questions if the number of years traded increases at a decreasing rate the MVQ estimates will increase. In the Years Gain questions if the amount of income traded increases at a decreasing rate we would expect the MVQ estimates to decrease. Finally, in the Years Loss questions, if the amount of income given up increases at an increasing rate we would expect the MVQ estimates to increase.

A further test of sensitivity to scale is whether the mean number of years (income) traded is influenced by the direction of the change. This gives rise to two further null hypotheses:

<u>Null Hypothesis 5</u>: *The mean number of years traded will not differ between the <u>Income Loss</u> and <u>Income Gain</u> questions*
See section 4.2.3.3.i for explanation.

<u>Null Hypothesis 6</u>: *The mean amount of income given up/accepted will not differ between the <u>Years Gain</u> and <u>Years Loss</u> questions*
Alternatively given diminishing marginal utility of income and life years we would expect the Years Loss questions to elicit considerably higher trades.

These two hypotheses are tested through both paired and unpaired t-tests. This particular sensitivity is also tested in relation to the MVQ estimates. Given the anticipated relationships outlined above we would expect the Income Gain questions to produce higher estimates than the Income Loss questions, and the Years Loss questions to produce higher estimates than the Years Gain questions.

6.2.3.4.ii Correlation with ability to pay

A further test of validity is whether the MVQ estimates are correlated with ability to pay.

Null Hypothesis 7: *The MVQ results will not differ depending on respondent income*

Alternatively, given the importance of respondent income in the MVQ calculation equations, the MVQ results are likely to be positively correlated with respondent income. Mean MVQs are compared by respondent income for the Income Loss, Income Gain and Years Loss questions. This analysis is not performed for the Years Loss questions due to the small sample size. Formal testing of this hypothesis is not possible due to the small sample sizes in the different income brackets.

6.2.3.4.iii Convergence with the ranking exercise

The results from the ranking exercise allow for some simple pairwise tests of consistency. Respondents who ranked 10 years with 80% of annual income as higher than 8 years with 100% of annual income should have traded less than 2 years in the 20% Income Loss questions. Visa versa, respondents who ranked 8 years with 100% of income as higher should have traded at least 2 years in the 20% Income Loss questions. A four box grid is created showing responses to the two different exercises. Two of the boxes represent consistent responses, while two represent inconsistent responses. The percentage of respondents giving consistent responses is then calculated. This analysis is conducted for the 20% Income Loss, 20% Income Gain and 2 Years Loss questions. No ranking was included for scenarios with duration beyond 10 years so the analysis can not be performed for the Years Gain questions.

6.2.3.5 The effect of background characteristics

Responses to the life satisfaction questions and the feedback questions are categorised as either 'yes' or 'no'. Mean number of years traded are then compared by whether respondents fall into the yes or no category. Comparison is through t-tests. Mean number of years traded is also compared with whether or not respondents' income exceeds their expectations.

Ordered logit regression analyses are performed to find determinants of the number of years traded. The data for number of years traded is discrete in nature, not normally distributed (see Appendix A. 6.2) and grouped around certain responses (e.g. zero). Therefore, estimation through standard OLS would have been inappropriate. For this reason the data was transformed into a categorical variable and modelled through ordered logit regressions. Separate regressions are performed for the Income Loss, Income Gain and Years Gain questions (regression analysis is not performed for the Years Loss questions due to the small sample sizes caused by the LG responses). Four regressions are performed for each question: one for each income/years change level, and one pooled model. The explanatory variables include background characteristics, responses to the feedback questions (as dummies), responses to the life satisfaction questions (as dummies) and dummies to represent the variant of questionnaire the respondent was faced with. In the pooled models dummies are included to show the effect of higher income/years change levels.

Initially regression models including a full set of explanatory variables were performed. Any variable that was not significant at the 5% level in any of the four models was then dropped to form reduced models. Wald tests were performed to test the null hypothesis that the variables that were dropped all simultaneously had coefficients equal to zero in the full models. Akaike's (1973) information criterion is presented as an indication of explanatory power (see section 4.2.3.2 for an explanation). Pregibon's (1981) link test is used to test for mis-specification. This test uses the linear predicted value and linear predicted value squared as the predictors to rebuild the model. If the model is properly specified the linear predicted value squared should have very little predictive power. Therefore, if it is significant this is indicative of mis-specification.

6.2.3.6 MVQ Results

MVQ estimates from the Income Loss and Income Gain questions are presented for the individual approach: LG responders are excluded and over-traders are truncated at zero. Results from these questions are then presented for the aggregate approach in which LG responders and over-traders are included. Estimates are also presented for the Years Loss and Years Loss questions based on both the individual and aggregate approaches. In both approaches LG responders can be included for the Years Gain questions, but are excluded for the Years Loss questions. Under-traders are truncated to zero in the individual approach but left unchanged in the aggregate approach. Further results using different sub-samples can be found in the Appendix.

6.3. Results

This section presents background characteristics of the sample, assess the feasibility, reliability and validity of the methods, and presents results produced by the methods.

6.3.1 Background Characteristics

Table 6.1 shows background characteristics for the full sample and, for comparison, the figures for the UK population. Details of sources of the UK figures can be found in the Appendix A6.3. The study sample has more females and a slightly higher mean age. Across the UK 72% are either employed or self-employed, which is similar to the figure of 72% in the study sample. Mean income also appears to be close to the national average. The religious makeup of the sample appears quite representative, although perhaps slightly less diverse as indicated by the lower percentage of people in the 'other' category. Looking at marital status it appears the study sample has fewer single people, but the UK statistics used did not have a category for co-habiting so these people will appear in the single category. There is less home ownership amongst the study sample than across the UK. Mean self-reported health on the EQ-VAS amongst the sample is very similar to the value obtained in the large scale Measurement and Valuation of Health Survey (Dolan *et al.* 1995, Kind *et al.* 1999) which elicited values from 3,381 members of the UK population. Appendix A6.4

Table 6.1 - Background Characteristics for the full sample and for the UK

		Full Sample	UK (Source in Appendix A6.3)
Number of Respondents		100	
Gender	Male	42%	49%
	Female	58%	51%
Age	Average	40.29	39
	SD	13.07	
	18-35	39%	
	36-50	38%	
	51-69	23%	
Educated beyond min school leaving age	Yes	85%	72%
	No	15%	28%
Educated to Degree Level	Yes	60%	31%
	No	40%	69%
Employment	Employed	54%	72%
	Self-Employed	24%	
	Retired	4%	
	Housework	4%	
	Student	5%	
	Seeking Work	6%	
	Other	3%	
Net Own Monthly Income	Mean	1653	1622
	<£430	14%	
	£430-£1300	33%	
	£1300-£2200	31%	
	>£2200	22%	
Children under 18 in the household	Yes	34%	29%
	No	66%	71%
Religion	Roman Catholic	11%	9%
	Protestant	43%	37%
	Atheist/Agnostic	39%	46%
	Other[1]	7%	19%
Marital Status	Married	35%	40%
	Single	31%	47%
	Co-habiting	25%	
	Divorced	7%	7%
	Widowed	2%	6%
Home Ownership[2]	Own/mortgage	54%	68%
	Rent	46%	32%
Mean Self-Reported Health on the EQ-VAS[3]		0.815	0.825

1. Other: Muslim (2), Jewish (1), Russian Orthodox (1), Spiritual (2), Not specified (1). Five respondents noted they were Christian, and five noted they were Church of England. These 10 respondents were entered as Protestant. 2. Relevant sample size=98. One respondent lived with their parents (V2) and one lived in co-op housing (V1). 3. Relevant sample size: 99. One respondent was unable to position dead on the scale (V4).

shows background characteristics by questionnaire variant. The Chi^2 tests show there were no significant differences between the four variants.

6.3.2 Feasibility

Only one respondents was unable to complete the task, giving a completion rate of 99%. The interviewer reported that this individual found all manner of aspects confusing and their responses could not be treated as reliable indicators of preference. This individual was a 59 year old male who received variant A of the questionnaire. This individual is excluded from all further analysis, reducing sample size to 99.

As in the previous chapter feasibility can be assessed by the prevalence of LG responses and over-trading behaviour. Table 6.2 shows the number of LG responses and over-traders/under-traders. The values in brackets show the number of respondents who gave LG responses/over-traded who also did so in the largest category. For example, of the 12 people who gave LG responses for 50% Income Loss, all 12 of them did so for 20% Income Loss. The results show that the number of LG responses (those who would not give up any time to maintain their income) falls as the level of Income Loss increases (significant at the 1% level). A total of nine respondents gave LG responses in all of the Income Loss questions.

The number of over-traders (those who give up a higher proportion of years than the proportion of income loss they are faced with) falls as the level of Income Loss increases (statistically insignificant). This is to be expected because as the level of income loss increases, respondents must trade progressively more time in order to be an over-trader. There is less consistency within respondents in Income Loss over-trading than LG responding. Of the 14 respondents who over-trade for 80% Income Loss, only four of them over-trade for 20% Income Loss. Only two respondents over-trade in all three Income Loss questions. A total of 42 respondents did not give LG responses or over-trade in any of the Income Loss questions.

More people give LG responses in the Income Gain questions than in the Income Loss questions. The number of non-traders falls as the level of gain increases (significant at the

1% level). There is strong within respondent consistency in LG responses in the Income Gain questions. All 19 respondents who gave LG responses for 80% gain, also did so for 50% gain and 20% gain. The number of over-traders falls as the level of Income Gain increases (significant at the 10% level). Only two respondents over-trade in all three Income Gain questions. A total of 38 respondents did not give LG responses or over-trade in any of the three Income Gain questions.

Table 6.2 - The number of LG responses and over/under trades (total n=99)

	LG responses	Over-traders/ Under-traders	Non-problematic responses
20% Income Loss	29	19	51
50% Income Loss	12 (12)	15 (10)	72
80% Income Loss	10 (10)	14 (4)	75
Chi^2 Test (p-value) [1]	**0.000**	0.593	**0.000**
LG responded/Over-traded in all 3 Income Loss questions	9	2	42
20% Income Gain	44	14	41
50% Income Gain	25 (25)	6 (4)	68
80% Income Gain	19 (19)	6 (4)	74
Chi^2 Test (p-value)	**0.000**	0.067	**0.000**
LG responded/Over-traded in all 3 Income Gain questions	19	2	38
2 Years Gain	12 (11)	8 (7)	91
5 Years Gain	15 (12)	19 (18)	80
8 Years Gain	17	30	69
Chi^2 Test (p-value)	0.602	**0.000**	**0.000**
LG responded/Under-traded in all 3 Years Gain questions	11	6	67
2 Years Loss	45 (45)	4 (2)	50
5 Years Loss	76 (75)	10	13
8 Years Loss	89	4 (2)	6
Chi^2 Test (p-value)	**0.000**	0.119	**0.000**
LG responded/Non-traded in all 3 Years Loss questions	44	1	3

1. Chi^2 test shows whether there is any significant relationship between being a LG responder or over/under trader and the level of income/years change level respondents are faced with.
The figures in brackets show the number of respondents that also LG respond or over/under trade in the largest category e.g. 50% income loss non-traders: 12 (12) means all 12 non-traders for 50% income loss were also non-traders for 20% income loss.

In the Years Gain questions, in order for an LG responder to maintain the same meaning i.e. someone who values life as infinitely more important than income, an LG responder is someone who would give up all of their income to achieve the increase in life years. As one would expect, as the number of Years Gained increases, the number of LG responses increases (statistically insignificant). A total of 11 respondents give LG responses in all three of the Years Gain questions. However, these do not represent problematic responses in these questions, because an LG response gives a value (100% of current annual income) that can be used to derive an MVQ estimate. Therefore, the Years Gain questions have the highest number of non-problematic responses. A total of 67 respondents do not give LG responses or under-trade in any of the three Years Gain questions. In the Years Gain questions the number of under-traders increases as the number of Years Gained increases (significant at the 1% level). One respondent would not give up any income in any of the three Years Gain questions to achieve an increase in life expectancy.

In the Years Loss questions being a LG responder means that a respondent will not accept any amount of income to compensate for a fall in life years. There are a lot of LG responders in these questions. LG responses become more prevalent as the number of years lost increases (significant at the 1% level). A total of 89 (out of 99) respondents would not accept any amount of income in the 8 Years Loss questions. There is strong consistency in LG responses in these questions. All 45 LG responders in the 2 Years Loss questions also give LG responses in the 8 Years Loss questions. There was no clear pattern in under-trading behaviour. In the 2 Years Loss questions 50 respondents did not give LG responses or over-trade, but only three respondents did not give LG responses or over-trade across all three Years Loss questions.

Comparing across the different types of questions, all nine of the respondents who gave LG responses in all of the Income Loss questions, also gave LG responses in all of the Income Gain questions and all of the Years Loss questions. Two of these nine gave LG responses in the Years Gain questions. Nine of the 11 respondents who gave LG responses in all three of the Years Gain questions also gave LG responses in all of the Years Loss questions.

The consistency in responses across the different questions in LG behaviour suggests that LG responses are a meaningful statement of preference for length of life over income. LG responders in the Income Loss and Income Gain are often prepared to give up all of their income to achieve an increase in life years in the Years Gain questions. If the LG responses in the Income Loss and Compensating Years Gain questions were protest responses we might expect these respondents to trade no income in the Years Gain questions.

Responses to the feedback questions also suggest LG responses and over/under trading behaviour may be meaningful. The mean response to the statement "I consider it unethical to give up years of life for an increase in income" was 2.62 where 1 is strongly agree and 5 is strongly disagree. The mean response to the statement "Time is more important than money because I want to be able to spend time with friends and family" was 1.61 which strongly supports the preference for length of life over income. With regards to over/under trading behaviour respondents typically did not try and calculate total lifetime income in the two scenarios (mean score 3.61), but the interviewers felt respondents typically understood the exercise well (mean score of 1.16, where 1 means they understood and 3 means they did not). Further exploration of any systematic relationships between LG responses and over/under trading behaviour and background and feedback questions may help to further determine whether these responses are meaningful.

6.3.2.1 Can LG responses be explained?

Bi-variate analysis (with Chi^2 tests) was performed to identify any significant relationships between background characteristics and the likelihood of being an LG responder in the four different TTO questions (see Appendix A6.5a-d). Relationships that are significant at the 1% level are discussed here. People in employment are less likely to be LG responders in both the Income Loss and Income Gain questions. This suggests that employed people value their income more highly. In the Years Loss questions Age has a highly significant effect on being an LG responder (would not accept any amount of money to compensate for fall in life years), with people in the youngest age bracket (18-35) being less likely to be LG responders. This suggests that younger people are not as motivated by money, perhaps because they do not yet have dependents or a mortgage. In the Years Loss questions home owners were more likely to be LG responders. The interview did not distinguish between owning the home outright or with a mortgage, making it difficult to make inferences. One explanation might be that home owners are financially more secure and hence they do not need to increase their income further by giving up life years.

Whether or not someone gave LG responses was compared against responses to the life satisfaction questions (see Appendix A6.6a-d). T-tests were performed comparing the mean response to the life satisfaction questions for those who gave LG responses and those who did not, for all four TTOs. There was only one point of significance. For the Years Loss questions, those who gave LG responses in all three questions were more likely to be satisfied with their income (significant at 5% level). Clearly respondents who are satisfied with their current income are unlikely to want to give up years of life to have a higher income.

Whether or not someone gave LG responses was compared against responses to the income expectations questions (see Appendix A6.7a-d). T-tests compared the mean responses to the four income expectations questions for those who did and did not give LG responses, for all four TTOs. There were no points of significance. Chi^2 tests also compared the proportions of respondents with an income that exceeded the four different income expectations levels, for those who did and did not give LG responses. There was only one point of significance. In the Income Gain questions a lower proportion of LG responders had an income that exceeded the level above which they would consider income to be excellent (significant at the 10% level). This finding is only weakly significant and contradicts the finding from the analysis of the life satisfaction questions for the Years Loss questions (above).

Further bi-variate analysis (with t-tests) was performed to identify any significant relationships between responses to the feedback questions and the likelihood of being an LG responder in the four different TTO questions (see Appendix A6.8a-d). Relationships that are significant at the 1% level are discussed here. Whether or not respondents attempted to calculate the total lifetime income they would have in each scenario had a significant effect on the likelihood of being an LG responder in the Income Loss, Income Gain and Years Gain questions. In all three cases LG responders were less likely to calculate lifetime income in each scenario. These respondents are likely to have a better understanding of the tasks and approached them from a more analytical, rather than philosophical, perspective. They are also likely to have given more thought to the exercise and may have been driven by a desire to give consistent answers or the 'right' answers. In the case of the most severe level of income loss (in the Income Loss questions) it is clear that respondents who did not want their income to drop below a certain level would chose to trade years of life. However, unless the threshold below which they do not want their income to fall is very high, it is not obvious why having in mind a minimum amount of income would influence trading. Perhaps, once again, responses to this question are indicative of respondents who have given more thought to the exercise. Respondents who thought it was unethical to give up years of life for an increase in income were more likely to give LG responses in the Years Gain questions (i.e. they are more likely to give up all of their income to achieve an increase in life years). This indicates a strong preference for life years over income.

6.3.2.2 Can over-trading/under-trading behaviour be explained?

Bi-variate analysis (with Chi^2 tests) was performed to identify any significant relationships between background characteristics and the likelihood of being an over/under-trader in the four different TTO questions (see Appendix A6.9a-d). Relationships that are significant at the 1% level are discussed here.

Home owners are less likely to be under-traders in the Years Loss questions. This is consistent with the analysis of LG responses presented earlier, and suggests home owners are financially secure and hence not interested in increasing their income further. Home owners were also less likely to be under-traders in the Years Gain questions. This suggests that not only are home owners not interested in increasing their incomes, they are also prepared to give up more income (than non-home owners) to maintain life years. Respondents with children were less likely to be under-traders in the Years Gain questions. This suggests parents want to be around to look after their children, even if this means forsaking a large amount of income.

Bi-variate analysis of responses to life satisfaction questions identified one variable that was significantly (1%) correlated with under-trading behaviour (see Appendix A6.10a-d). Respondents who were satisfied with their income were less likely to be under-traders in the Years Loss and Years Gain questions. This confirms the findings in the analysis of LG responses, and suggests financially secure respondents are not interested in increasing their income further at a cost to their life years and are also prepared to give up more income to maintain their life years.

Income expectations were found to be significantly correlated (1%) with under-trading behaviour in the Years Loss and Years Gain questions (see Appendix A6.11a-d). Under-traders in the Years Loss questions had lower bounds of income sufficiency, and a lower level above which income would be considered excellent. A small percentage of the under-traders in the Years Gain questions have an income they consider to exceed the bounds of income sufficiency. This further confirms the finding that wealthier respondents (from a self-perceived perspective) are less interested in increasing their income at a cost to their life years, and are also more prepared to give up income to maintain life years, which is consistent with the theory of diminishing marginal utility of income.

Responses to one feedback question were significantly (1%) correlated with under/over-trading behaviour (see Appendix 6.12a-d). Respondents who wanted to be able to spend time with their family and friends were less likely to be over-traders in the Income Loss questions. This suggests these people are prepared to forsake income to achieve their aim.

6.3.2.3 Summary of Feasibility

The completion rate was high at 99%, suggesting that the respondents were able to understand the tasks they were faced with.

LG responses can either represent a meaningful preference for length of life over income, or protest responses where the values given do not meaningfully represent their preference. The consistency of LG responses across the different types of question (e.g. giving up no years in the Income Loss questions and giving up all of their income in the Years Gain questions) suggest respondents are displaying a meaningful preference for life over income. Statistically significant relationships between LG behaviour and background and feedback questions, which are typically intuitive, further suggest this is meaningful behaviour at the aggregate level.

Over/under trades can either represent respondents not understanding the tasks or a meaningful preference for high annual income over length of life. Although it appears that respondents did not attempt to calculate lifetime income in each scenario, the interviewers reported that respondents seemed to understand the tasks well. There were also a number of statistically significant variables that suggest over/under trading behaviour was, on the whole, not caused by a lack of understanding.

6.3.3 Reliability

Obviously the data available do not allow for the preferred test-retest reliability analysis. However, internal reliability can be assessed by observing whether logical consistency is adhered to (as detailed in the methods section). Table 6.3 shows the number of consistent responses for the Income Loss, Income Gain and Years Gain questions. The analysis was not performed for the Years Loss questions due to the very high number of LG responses. The results show that the Income Loss questions clearly elicit the most consistent responses. Of 99 respondents, 55 valued the three income change levels in a manner consistent with logical expectations: that is they traded more years to avoid a 50% income loss than they did to avoid a 20% income loss, and they traded yet more years to avoid an 80% income loss. A further 23 respondents were weakly consistent (i.e. two of the three income levels were valued as equal), and only 12 gave inconsistent responses. The Income Gain questions gave few consistent responses but only slightly more inconsistent responses. However, in the Years Gain questions over a third of respondents gave inconsistent responses.

Table 6.3 - Test of Logical Consistency			
	Income Loss	Income Gain	Years Gain
Total LG responder	9	19	11
Consistent	55	29	24
Weak Consistent	23	35	30
Inconsistent	12	16	34

6.3.4 Validity

As in the previous chapter internal validity is assessed through sensitivity to scale and correlation with ability to pay. In this chapter a simple test of convergent validity is also performed using the results from the ranking exercise.

6.3.4.1 Sensitivity to Scale

Sensitivity to scale is tested both in terms of the number of years traded and the subsequent MVQ estimates that are produced. Furthermore, as in the previous chapter sensitivity is assessed in relation to the size of the income/years change level and the direction of the change.

6.3.4.1.i Sensitivity of the number of years (or amount of income) traded to income (years) change level

Table 6.4 shows the mean number of years given up in the Income Loss and Income Gain questions, including LG responders. The results excluding LG responses for the Income Loss and Income Gain questions can be found in the Appendix (Table A6.13). Table 6.5 shows the percentage increase in income required to compensate for a fall in life expectancy in the Years Loss questions, and the percentage of income that would be given up to achieve an increase in life expectancy in the Years Gain questions. The Years Gain values include LG responses, but the Years Loss values do not since LG responses in these questions cannot be allocated a value (their responses can only be interpreted qualitatively).

In the 2 Years Loss and 8 Years Loss questions the mean amount of income given up is very high. Both values are influenced by one outlier of 1 million percent (i.e. 999,900% increase). If these values are excluded the value for 2 Years Loss is 307%, and the value for 8 Years Loss is 901%. However, there is no strong grounds for excluding these outliers which is why they have been included in Table 6.5.

In the Income Loss and Income Gain questions the mean number of years given up increases as the level of income change increases. Paired t-tests were performed between each of the income change levels for each question. For example, in the Income Loss questions paired t-tests were performed between 20% income loss and 50% income loss, between 50% income loss and 80% income loss, and between 20% income loss and 80% income loss. The differences between the income change levels were all significant at the 1% level for both the Income Loss and Income Gain.

The results of the Years Loss questions in Table 6.5 show that the amount of income required to compensate for a fall in life years increases as the size of the fall in years increases. However, none of these differences are significant.

The results of the Years Gain questions in Table 6.5 show that the amount of income given up to achieve an increase in life years increases as the number of life years to be gained increases. The differences between 2 Years Gain and 5 Years Gain, and between 2 Years Gain and 8 Years Gain, are both significant at the 1% level. The differences between 5 Years Gain and 8 Years Gain are not significant.

These results suggest that the questions, excluding Years Loss, are sensitive to scale. This finding is further supported by the results of the pooled regression analysis presented in section 6.3.6. The level of income loss in the Income Loss and Income Gain questions has a highly significant effect on the number of years traded. Likewise, in the Years Gain the number of years gained has a highly significant effect on the amount of income traded.

Table 6.4 - Number of years traded in the Income Loss and Income Gain questions

		Income Loss			Income Gain		
		20%	**50%**	**80%**	**20%**	**50%**	**80%**
Number of years traded to either avoid an income loss or achieve and an income gain <u>including</u> LG responses	N	99	99	99	99	99	99
	Mean	1.10	2.42	4.35	0.69	1.14	1.52
	SD	1.47	2.41	3.17	1.32	1.27	1.54
	10th Percentile	0.00	0.00	0.00	0.00	0.00	0.00
	Median	0.50	1.75	3.75	0.17	0.50	1.25
	90th Percentile	2.75	6.00	9.25	2.50	2.00	3.50
	Paired t-test loss vs gain (p value)	**0.007**	**0.000**	**0.000**			
	Paired t-test	20% Loss vs 50% Loss	50% Loss vs 80% Loss	20% Loss vs 80% Loss	20% Gain vs 50% Gain	50% Gain vs 80% Gain	20% Gain vs 80% Gain
	p-value	**0.000**	**0.000**	**0.000**	**0.002**	**0.000**	**0.000**

Table 6.5 – Amount of income traded in the Years Gain and Years Loss questions

		Years Loss[1]			Years Gain		
		2 yrs	**5 yrs**	**8 yrs**	**2 yrs**	**5 yrs**	**8 yrs**
% of income either required to compensate for a loss in years, or given up to achieve an increase in years	N	54	23	10	99	99	99
	Mean	18,818	843	100,801	51.8	61.2	62.8
	SD	136,034	2211	315,915	30.1	28.1	29.1
	10th Percentile	25	65	80	20.0	22.5	22.5
	Median	66	100	400	45.0	45.0	65.0
	90th Percentile	400	900	4900	100.0	100.0	100.0
	Unpaired t-tests Loss vs Gain (p-value)	0.171	**0.001**	**0.001**			
	T-test (Unpaired for Loss, paired for Gain)	2 yrs Loss vs 5 yrs Loss	5 yrs Loss vs 8 yrs Loss	2 yrs Loss vs 8 yrs Loss	2 yrs Gain vs 5 yrs Gain	5 yrs Gain vs 8 yrs Gain	2 yrs Gain vs 8 yrs Gain
	p-value	0.530	0.131	0.176	**0.000**	0.252	**0.000**

1. Excluding LG responses

6.3.4.1.ii Sensitivity of the number of years traded to the direction of income change

More years are given up in the Income Loss questions than the Income Gain questions, for all three income change levels (significant at 1%). This supports the theory of diminishing marginal utility of income.

The vast difference between the amount of income given up in the Years Loss questions, and the amount of income required in the Years Gain questions is likely to be caused by a combination of both the budget constraint and diminishing marginal utility of life years.

6.3.4.1.iii Sensitivity of the MVQ estimates to the income/years change levels

Table 6.6 shows mean MVQ values for the Income Loss and Income Gain questions calculated at the individual level, excluding LG responses. Over-traders are included by truncating negative values at zero. Some very high individual values were produced because as the number of years traded becomes very small the size of the MVQ result increases rapidly (smallest unit of trade was one week). A respondent with an annual income of £10,000 who traded one week in the 20% Income Loss question, would generate an MVQ value of £990,000. As with the previous chapter results are presented using individual income and household income. Results with a balanced panel, and excluding over-traders are presented in the appendix (Tables A6.14 and A6.15).

The Income Loss results decrease as the level of loss increases. However, these differences were only significant (5% level) when comparing the 80% loss result with the 20% loss result. The Income Gain results increase as the level of gain increases, but these differences were not significant. This shows that while the method is sensitive to scale in terms of the number of years traded, this does not lead to statistically significant differences in MVQ estimates. This is a positive finding that suggests the MVQ estimates are fairly robust to the level of income change specified.

Table 6.6 - MVQ values calculated at the individual level using the Income Loss and Income Gain questions (*excluding* LG responses) based on both individual income and household income

		Income Loss			Income Gain		
		20%	**50%**	**80%**	**20%**	**50%**	**80%**
		(n=70)	(n=86)	(n=88)	(n=55)	(n=74)	(n=79)
Mean number of years traded		1.55	2.76	4.84	1.24	1.52	1.88
Mean Annual Individual Income (£)		19,220	19,191	19,034	21,159	20,061	19,559
Mean Annual Household Income (£)		33,554	32,046	31,486	36,231	34,131	33,144
Mean VAS Own Health		0.817	0.821	0.820	0.790	0.814	0.819
Number of negative responses (truncated at zero)		18	15	13	14	6	6
Value of 1 Quality Adjusted Life Year (Based on Individual Income, £)	Mean	**91,418**	**80,160**	**70,404**	**166,644**	**269,880**	**268,000**
	SD	210,087	125,786	167,371	435,171	634,496	601462
	10th Percentile	0	0	0	0	927	1,673
	Median	14,617	26,224	10,651	29,568	44,131	54,353
	90th Percentile	245,435	296,400	218,400	382,500	634,496	585,900
	Paired t-test Loss vs Gain (p-value)	0.446	**0.006**	**0.002**			
	Paired t-test	20% Loss vs 50% Loss	50% Loss vs 80% Loss	20% Loss vs 80% Loss	20% Gain vs 50% Gain	50% Gain vs 80% Gain	20% Gain vs 80% Gain
	p-value	0.227	0.521	**0.034**	0.059	0.687	0.239
Value of 1 Quality Adjusted Life Year (Based on Household Income, £)	Mean	**166,401**	**129,501**	**107,730**	**227,547**	**453,564**	**523,178**
	SD	382,423	195,441	223,355	481,738	951,354	1,201,139
	10th Percentile	0	0	0	0	1,855	6,233
	Median	25,204	47,369	17,780	43,958	80,434	108,706
	90th Percentile	420,713	448,800	405,000	694,832	1,405,102	2,247,243

Table 6.7 shows MVQ results calculated at the aggregate level, including both LG responses and over-traders. This approach allows the use of the full sample. The results are consistently lower than those in Table 6.6 as the effect of very small trades is eliminated. The Income Loss questions again produce fairly consistent results while the Income Gain questions are more variable. The aggregate approach does not allow for statistical testing as in the individual approach.

Table 6.8 shows MVQ results based on the Years Gain and Years Loss questions. Negative responses generated by under-traders are truncated to zero. LG responses in the Years Loss questions are excluded. LG responses in the Years Gain questions are respondents who give up all of their income and can be included in the analysis. Since the sample sizes for these questions are very small the very high values (truncated to £1 million) have a large impact on the means.

The extreme values in the Years Loss questions make it difficult to draw meaningful conclusions from the data. There were no significant differences in values generated by the different Years Loss levels. Considering the Years Gain questions the results fall as the number of years to be gained increases and the differences are all significant. This could partly be the effect of diminishing marginal utility of life years, but is more likely to be a product of the restrictive nature of the calculation method, in which respondent income acts as a considerable budget constraint. This also explains why the SDs, medians and 90th percentiles also fall as the number of years to be gained increases. Table 6.9 shows the MVQ estimates based on the Years Gain and Years Loss questions using the aggregate calculation method. The results generated by the Years Gain questions again fall as the number of years gained increases, while there is no clear relationship for the Years Loss questions.

Table 6.7 - MVQ values calculated at the aggregate level using the Income Loss and Income Gain questions, including LG responses and over-traders

	Equivalent Loss			Compensating Gain		
	20%	**50%**	**80%**	**20%**	**50%**	**80%**
Number of respondents	98	98	98	98	98	98
Mean number of years traded	1.11	2.43	4.36	0.70	1.15	1.53
Mean Annual Individual Income (£)	20,133	20,133	20,133	20,133	20,133	20,133
Mean Annual Household Income (£)	33,227	33,227	33,227	33,227	33,227	33,227
Mean VAS of Own Health	0.814	0.814	0.814	0.814	0.814	0.814
Value of a QALY (Individual Income)	**19,911**	**26,137**	**20,607**	**41,290**	**70,717**	**85,143**
Value of a QALY (Household Income)	**32,861**	**43,136**	**34,010**	**68,145**	**116,710**	**140,518**

Table 6.8 - MVQ values calculated at the individual level using the Years Gain and Years Loss questions, based on both individual income and household income

		Years Loss (Excluding LG responses)			Years Gain (Including LG responses)		
		2 yrs	**5 yrs**	**8 yrs**	**2 yrs**	**5 yrs**	**8 yrs**
Number of Respondents		54	23	10	98	98	98
Mean income given up/required		18,818	843	100,801	52	61	63
Mean Annual Individual Income (£)		17,201	16,877	14,024	20,133	20,133	20,133
Mean Annual Household Income (£)		29,445	23,732	18,248	33,227	33,227	33,227
Mean VAS Own Health		0.809	0.725	0.713	0.814	0.814	0.814
Number of negative responses (truncated at zero)		4	10	4	8	19	30
Value of 1 Quality Adjusted Life Year (Based on Individual Income, £)	Mean	**3,922,435**	**384,436**	**4,442,244**	**62,451**	**26,228**	**15,502**
	SD	26,800,000	284,044	13,900,000	93,317	38,174	23,749
	10th percentile	0	0	0	0	0	0
	Median	30,026	0	4,714	24,445	12,948	5,301
	90th Percentile	550,588	225,000	22,200,000	164,211	66,122	38,716
	Unpaired t-test Loss vs Gain (p-value)	0.155	**0.010**	**0.013**			
	T-tests (Unpaired for Loss, paired for Gain)	2yrs Loss vs 5 yrs Loss	5yrs Loss vs 8 yrs Loss	2 yrs Loss vs 8 yrs Loss	2 yrs Gain vs 5 yrs Gain	5 yrs Gain vs 8 yrs Gain	2 yrs Gain vs 8 yrs Gain
	p-value	0.530	0.167	0.953	**0.000**	**0.000**	**0.000**
Value of 1 Quality Adjusted Life Year (Based on Household Income, £)	Mean	**4,157,230**	**417,197**	**145,910**	**101,317**	**43,603**	**25,976**
	SD	26,800,00	1,360,368	322,945	166,092	69,029	43,446
	10th percentile	0	0	0	0	0	0
	Median	40,237	0	4,714	44,200	21,472	10,135
	90th Percentile	745,534	445,440	386,591	275,625	112,500	70,313

Table 6.9 - MVQ values calculated at the aggregate level using the Years Gain and Years Loss questions, based on both individual income and household income

	Years Loss (Excluding LG responses)			Years Gain (Including LG responses)		
	2 yrs	**5 yrs**	**8 yrs**	**2 yrs**	**5 yrs**	**8 yrs**
Number of Respondents	52	21	9	98	98	98
Mean number of years traded	122	219	901	52	61	63
Mean Annual Individual Income (£)	17,393	15,341	14,409	20,133	20,133	20,133
Mean Annual Household Income (£)	29,703	22,850	17,929	33,227	33,227	33,227
Mean VAS of Own Health	0.810	0.732	0.726	0.814	0.814	0.814
Value of a QALY (Individual Income)	**83,485**	**24,850**	**24,865**	**52,538**	**52,156**	**41,260**
Value of a QALY (Household Income)	**142,577**	**37,012**	**30,939**	**86,708**	**86,077**	**68,094**

6.3.4.1.iv Sensitivity of the MVQ estimates to the direction of income change

Looking at the MVQ results in Table 6.6, based on individual income, the Income Gain questions produce higher results than the Income Loss questions for all three income change levels. Paired t-tests showed that these differences were significant at the 1% level for the 50% and 80% income change levels, but insignificant for the 20% change level. Using a balanced panel (see Appendix A6.15) has no effect on the trends in the results.

Table 6.8 shows that the Years Loss questions consistently produce higher MVQ results than the Years Loss questions. These differences were significant at the 5% level for the 5 years and 8 years change levels. This finding is consistent with the finding in the contingent valuation literature that the WTA method typically elicits larger values than the WTP method.

6.3.4.2 Correlation with ability to pay

Tables 6.10a-6.10c show mean MVQs by respondent income level for the Income Loss, Income Gain and Years Gain questions. Results are presented for the individual approach and the aggregate approach. This analysis was not performed for the Years Loss questions because the sample size in the different categories would be too small. Table 6.10a shows that in the Income Loss questions as respondent income increases mean MVQ results increase at an increasing rate. This holds for both the individual approach and the aggregate approach. Table 6.10b shows that this relationship also holds in the Income Gain questions when using the aggregate approach. In the individual approach the relationship is not perfect as the second income bracket produces the lowest results. Finally, Table 6.10c shows that the relationship also holds for the Years Gain questions using both the individual and aggregate approaches. While from an ethical point of view dependence upon ability to pay may be seen as undesirable, from a methodological perspective one would expect values to be correlated with ability to pay as a test of validity. The results in Tables 6.10a-6.10c suggest that the Income Loss, Income Gain and Years Gain questions all correlate with ability to pay.

Table 6.10a - Weighted mean QALY values for different income brackets for both the individual approach and the aggregate approach, based on the Income Loss questions

	Individual Income Level		20%	n	50%	n	80%	n	Weighted Mean MVQ (£)
Individual Approach[1]	<£430 per month	Mean Years Traded	1.11		1.87		4.33		
		Mean VAS Own Health	0.96		0.96		0.94		
		MVQ	**27,052**	8	**36,058**	11	**34,748**	12	**33,227**
	£430 - £1300	Mean Years Traded	1.99		3.30		5.38		
		Mean VAS Own Health	0.77		0.78		0.78		
		MVQ	**58,357**	27	**27,631**	29	**49,555**	29	**44,871**
	£1300 - £2200	Mean Years Traded	1.54		3.01		5.14		
		Mean VAS Own Health	0.79		0.79		0.79		
		MVQ	**112,130**	22	**74,993**	29	**75,623**	30	**85,313**
	>£2200	Mean Years Traded	0.94		2.04		3.84		
		Mean VAS Own Health	0.87		0.86		0.86		
		MVQ	**163,316**	13	**207,123**	17	**112,422**	17	**160,753**
		Total N		**70**		**86**		**88**	
Aggregate Approach[2]	<£430 per month	Mean Years Traded	0.68		1.58		3.99		
		Mean VAS Own Health	0.94		0.94		0.94		
		MVQ	**6,010**	13	**6,740**	13	**3,129**	13	**5,293**
	£430 - £1300	Mean Years Traded	1.68		2.99		4.88		
		Mean VAS Own Health	0.79		0.79		0.79		
		MVQ	**2,597**	32	**9,017**	32	**8,590**	32	**6,735**
	£1300 - £2200	Mean Years Traded	1.09		2.81		4.98		
		Mean VAS Own Health	0.77		0.77		0.77		
		MVQ	**22,606**	31	**21,077**	31	**16,471**	31	**20,051**
	>£2200	Mean Years Traded	0.56		1.58		2.97		
		Mean VAS Own Health	0.84		0.84		0.84		
		MVQ	**132,488**	22	**111,210**	22	**86,612**	22	**110,103**
				98		**98**		**98**	

1. Excluding LG responses and including over-traders (truncated to zero)
2. Including LG responses ad over-traders

Table 6.10b - Weighted mean QALY values for different income brackets for both the individual approach and the aggregate approach, based on the <u>Income Gain</u> questions

	Individual Income Level		20%	n	50%	n	80%	n	Weighted Mean MVQ (£)
Individual Approach[1]	<£430 per month	Mean Years Traded	1.58		1.14		1.20		
		Mean VAS Own Health	0.99		0.98		0.96		
		MVQ	**20,910**	4	**75,794**	8	**195,884**	10	**120,402**
	£430 - £1300	Mean Years Traded	1.31		2.01		2.45		
		Mean VAS Own Health	0.72		0.77		0.77		
		MVQ	**96,658**	19	**145,996**	25	**74,941**	26	**106,212**
	£1300 - £2200	Mean Years Traded	0.96		1.48		1.87		
		Mean VAS Own Health	0.79		0.79		0.79		
		MVQ	**96,508**	21	**127,290**	26	**185,673**	27	**139,856**
	>£2200	Mean Years Traded	1.54		0.97		1.47		
		Mean VAS Own Health	0.84		0.85		0.85		
		MVQ	**273,420**	11	**380,760**	15	**369,040**	16	**348,182**
				55		**74**		**79**	
Aggregate Approach[2]	<£430 per month	Mean Years Traded	0.49		0.70		0.93		
		Mean VAS Own Health	0.94		0.94		0.94		
		MVQ	**9,060**	13	**17,651**	13	**21,345**	13	**16,019**
	£430 - £1300	Mean Years Traded	0.78		1.57		1.99		
		Mean VAS Own Health	0.79		0.79		0.79		
		MVQ	**18,371**	32	**22,634**	32	**29,909**	32	**23,638**
	£1300 - £2200	Mean Years Traded	0.65		1.24		1.63		
		Mean VAS Own Health	0.77		0.77		0.77		
		MVQ	**51,182**	31	**68,529**	31	**84,652**	31	**68,121**
	>£2200	Mean Years Traded	0.77		0.66		1.07		
		Mean VAS Own Health	0.84		0.84		0.84		
		MVQ	**71,291**	22	**308,964**	22	**290,306**	22	**223,520**
				98		**98**		**98**	

1. Excluding LG responses and including over-traders (truncated to zero)
2. Including LG responses and over-traders

Table 6.10c - Weighted mean QALY values for different income brackets for both the individual approach and the aggregate approach, based on the Years Gain questions

	Individual Income Level		2 yr gain	n	5 yr gain	n	8 yr gain	n	Weighted MVQ (£)
Individual Approach[1]	<£430 per month	Mean Years Traded	62		68		67		
		Mean VAS Own Health	0.93		0.93		0.93		
		MVQ	**10,438**	13	**4,017**	13	**2,576**	13	**5,677**
	£430 - £1300	Mean Years Traded	43		55		57		
		Mean VAS Own Health	0.79		0.79		0.79		
		MVQ	**20,756**	32	**9,297**	32	**5,764**	32	**11,939**
	£1300 - £2200	Mean Years Traded	48		57		62		
		Mean VAS Own Health	0.77		0.77		0.77		
		MVQ	**63,952**	31	**26,536**	31	**17,087**	31	**35,858**
	>£2200	Mean Years Traded	65		73		71		
		Mean VAS Own Health	0.84		0.84		0.84		
		MVQ	**151,718**	22	**63,548**	22	**35,071**	22	**83,446**
				98		**98**		**98**	
Aggregate Approach[2]	<£430 per month	Mean Years Traded	63		68		67		
		Mean VAS Own Health	0.94		0.94		0.94		
		MVQ	**8,642**	13	**8,122**	13	**6,462**	13	**7,742**
	£430 - £1300	Mean Years Traded	43		55		57		
		Mean VAS Own Health	0.79		0.79		0.79		
		MVQ	**21,268**	32	**22,084**	32	**15,351**	32	**19,568**
	£1300 - £2200	Mean Years Traded	48		57		62		
		Mean VAS Own Health	0.77		0.77		0.77		
		MVQ	**50,584**	31	**47,691**	31	**41,916**	31	**46,730**
	>£2200	Mean Years Traded	65		73		71		
		Mean VAS Own Health	0.84		0.84		0.84		
		MVQ	**147,970**	22	**152,718**	22	**121,121**	22	**140,603**
				98		**98**		**98**	

1. Including LG responses and under-traders (truncated to zero)
2. Including LG responses and under-traders

6.3.4.3 Convergent Validity

Results from the ranking exercise allow a simple test of convergent validity to be performed, as specified in the methods section. In order to pass this test respondents who ranked 10 years with 80% of annual income as preferable to 8 years with 100% of annual income should have traded less than 2 years in the 20% Income Loss questions. Vica Versa, respondents who ranked 8 years with 100% of annual income as preferable to 10 years with 80% of annual income should have traded 2 or more years in the 20% Income Loss questions. Similar tests are performed for the Income Gain and Years Loss questions. No scenario with life years beyond ten was included in the ranking exercise so no equivalent test can be performed on the Years Gain questions.

The results in Table 6.11 show that 74% of responses to the 20% Income Loss questions were consistent with the ranking exercise. In the 20% Income Gain questions 88% of responses were consistent with the ranking exercise, while in the Years Loss questions 92% of responses were consistent.

Table 6.11- Test of convergence with responses to the ranking exercises

	Income Loss			
	Traded ≥ 2 years	Traded <2 years		
Ranked 10 years with 80% of income>8 years with 100% of income	13	67		
Ranked 10 years with 80% of income<8 years with 100% of income	9	10		
Total Consistent Responses	76			
Total Inconsistent Responses	23			
Percentage Consistent	77%			
	Income Gain		Years Loss	
	Traded ≥ 2 years	Traded <2 years	Accepted ≤20% income gain	Would only accept more than 20% income gain
Ranked 10 years with 100% of income >8 years with 120% of income	7	82	1	88
Ranked 10 years with 100% of income <8 years with 120% of income	5	5	3	7
Total Consistent Responses	87		91	
Total Inconsistent Responses	12		8	
Percentage Consistent	88%		92%	

Note: Shaded cells represent inconsistent responses

6.3.5 The effect of background characteristics

Bi-variate analysis showing the relationships between responses to the life satisfaction questions and the mean number of years traded, and between responses to the feedback questions and the mean number of years traded, can be found in the Appendices A6.16a-A6.17c. Analysis is not performed on the Years Loss questions due to the high number of LG responses.

Tables 6.12a and 6.12b shows reduced regression models showing the effect of background characteristics, life satisfaction and response to the feedback questions on the number of years traded in the Income Loss and Income Gain questions respectively. Due to the discrete nature of the data, the dependent variables (number of years traded) were transformed into categorical variables and ordered logit regressions were performed. One model is estimated for each income change level and a pooled model is also estimated.

Initially full models (see Appendix A6.18a-b) were estimated. In order to generate the reduced models any variable that was not significant in any of the four full models (at 5% level) was dropped. Wald tests were performed to test the null hypothesis that the variables that were dropped all simultaneously had coefficients equal to zero in the full models. The Wald statistics all suggest there was no loss of explanatory power as a result of removing the insignificant variables. The Link tests are not indicative of mis-specification in any of the models and the AIC values show improved explanatory power in the reduced models compared to the full models.

The results in Table 6.12a show that there is a significant U-shaped age relationship in all four models (as seen in TTO for health). This relationship also holds for three of the four Income Gain models in Table 6.12b. In all cases the bottom of the curve occurs at a higher age than the oldest member of the sample, meaning that increased age always has a negative effect on the number of years traded, but at a diminishing rate. This is consistent with the findings in the bivariate analysis presented earlier. A possible explanation for this finding is that older respondents have fewer years of life left to live and so value each year more highly.

Table 6.12a - The effect of background characteristics, life satisfaction and feedback on the number of years traded in the Income Loss questions (ordered logit)

	20% Income loss	50% income loss	80% income loss	Pooled (Long Dataset)
Age	**-0.254****	**-0.269*****	**-0.246****	**-0.207****
Age^2	**0.003****	**0.003*****	**0.003****	**0.002****
Atheist/Agnostic=0, Religious=1	**-1.250*****	**-0.910****	-0.504	**-0.968*****
Employed=1, Not employed=0	**1.593****	0.585	0.266	0.727
Thought about family and friends	**-1.219*****	**-1.528*****	**-1.423*****	**-1.345*****
Aid was helpful	-0.688	**-0.728***	**-0.808***	**-0.670***
Attempted to calculate lifetime income	**1.057****	**0.791****	**0.923****	**0.914*****
I am satisfied with my income	**-1.037*****	**-1.195*****	-0.422	**-0.836****
50% income loss vs 20% income loss				**1.458*****
80% income loss vs 20% income loss				**2.909*****
Number of Observations	99	99	99	297
Pseudo R^2	0.157	0.162	0.126	0.211
Link test	0.231	0.864	0.131	0.169
Wald Test	0.944	0.531	0.326	0.571
Akaike Information Criterion	2.539	2.409	2.520	2.235

Values presented are coefficients
* Significant at 10% level ** Significant at 5% level *** Significant at 1% level
Dependent Variable categories: **20% Income Loss**: 0 yrs traded=0; 0-0.5 yrs traded=1; 0.5-2 yrs traded=2; .2 yrs traded=3.
50% Income Loss: 0yrs traded=0; 0-2 yrs traded=1; 2-3 yrs traded=2; >3 yrs traded=3.
80% Income Loss: 0 yrs traded=0; 0-3 yrs traded=1; 3-6yrs traded=2; >6 yrs traded=3
Pooled: 0 yrs traded=0; 0-1 yrs traded=1; 1-3 yrs traded=2; >3 yrs traded=3.

Being religious had a negative effect on the number of years traded in three of the four Income Loss models. It could be that religious respondents are more likely to give LG responses, but this relationship was not found in the bivariate analysis presented earlier and whether or not respondents thought it was unethical to give up years of life for income was not found to be significant in the regression analysis. Being religious was not a significant explanatory variable in the Income Gain regression models.

Being in employment had a significant positive effect on the number of years traded in the 20% Income Loss model and three of the Income Gain models. This might be the case because a given percentage of income is likely to represent a larger amount for employed people than for unemployed people. However, if this explanation was valid we might expect income to reveal itself as significant explanatory variable. Alternatively, perhaps a certain amount of unemployment is voluntary, indicating that this group attach less value to income. Whether or not someone thought about their family and friends was significant in all Income Gain and all Income Loss models and had a negative effect on the number of years traded. It is clear that people who valued time with their friends and family, and had this in the forefront of their mind, traded fewer years than those who either did not value time with their family and friends or did not think about it.

Interestingly, attempting to calculate total lifetime income was significant in all Income Loss and all Income Gain questions. Respondents who attempted this traded more life years on average. Both the Income Loss and Income Gain pooled models show that the questions are sensitive to the size of the income loss/gain that respondents are faced with.

Table 6.12b - The effect of background characteristics, life satisfaction and feedback on the number of years traded in the Income Gain questions (Ordered Logit)

	20% Income gain	50% income gain	80% income gain	Pooled (Long Dataset)
Age	-0.202*	**-0.234****	**-0.329*****	**-0.215****
Age2	0.002	**0.003****	**0.004*****	**0.002****
Employed=1, Not employed=0	**2.241*****	0.968	**1.115***	**1.354****
I can rely on my partner's income	0.499	**0.732***	**0.836****	**0.676***
Thought about family and friends	**-1.285*****	**-1.681*****	**-1.541*****	**-1.561*****
Had in mind a minimum amount I could survive on	-0.258	-0.497	**-0.877****	**-0.628***
Attempted to calculate lifetime income	**1.584*****	**1.403*****	**1.723*****	**1.455*****
I am satisfied with my health	-0.703	-0.577	**-0.547***	-0.594
50% income gain vs 20% income gain				**1.039*****
80% income gain vs 20% income gain				**1.604*****
Number of Observations	99	99	99	297
Pseudo R^2	0.173	0.134	0.155	0.162
Link Test	0.488	0.956	0.269	0.215
Akaike Information Criterion	1.906	2.611	2.549	2.389

Values Presented are coefficients
* Significant at 10% level ** Significant at 5% level *** Significant at 1% level
Dependent variable categories: 20% Income Gain: 0 yrs traded=0; 0-1 yrs traded=1; >1yr traded =2
50% Income Gain: 0 yrs traded=0; 0-0.5 yrs traded=1; 0.5-1.5 yrs traded=2; >1.5 yrs=3
80% Income Gain: 0 yrs traded=0; 0-1 yrs traded=1; 1-2 yrs traded =2; >2yrs=3
Pooled: 0 yrs traded=0; 0-0.5 yrs traded=1; 0.5-2 yrs traded=2; >2 yrs traded=3

Analysis of the Years Gain questions revealed very few significant variables, and the estimation of reduced models proved problematic removal of insignificant variables produced significant Wald tests. The full models for the Years Gain questions are presented in the Appendix (A6.18c). Respondents who thought about their family and friends, on average, traded more income to gain years i.e. valued life years more highly which is consistent with the findings in the Income Loss and Income Gain models above. Respondents who attempted to calculate total lifetime income traded on average less income to gain life years i.e. attached less value to life years which is again consistent with the findings in the Income Loss and Income Gain models above. Finally, the pooled model showed that the Years Gain questions were also sensitive to the size of the outcome being valued.

Regression analysis was not performed on the Years Loss questions due to the very large number of LG responses.

6.3.6 The MVQ Results

Table 6.5 shows that the MVQ results derived from the Income Loss questions are fairly consistent and range from £70K to £91K. However, these values exclude LG responders. The evidence seems to suggest that LG responses may be meaningful statements of preference for length of life over income, and hence they should be included in the analysis. To do this the aggregate calculation method must be used. Using this approach the Income Loss questions produce values ranging from £20K to £26K. However, these values dilute the effect of the very small trades and represent a movement away from the principles of welfare economics. The Income Gain questions, when analysed using the individual approach produce estimates ranging from £166K to £276K. Using the aggregate approach produces estimates ranging from £41K to £85K.

The Years Loss questions elicit extreme responses and hence produce MVQ estimates ranging from £384K to over £4 million. Furthermore, use of the aggregate approach is not able to deal with LG respnders. The Years Gain questions produce very consistent MVQ estimates, ranging from £15K to £62 using the individual approach, and from £41K to £52K. However, this is largely driven by the strong budget constraint.

6.4 Discussion

6.4.1 Summary of Results

This chapter has presented a UK based interview study with a sample of 100 members of the general public to test four different TTO based questions to elicit an MVQ. This section summarises the results.

6.4.1.1 Feasibility

LG responding and over/under trading still represent a significant problem in all questions. The Years Gain questions produced the most usable results (69%-91% across the three questions) but these must be viewed with caution due to the influence of the budget constraint. The Years Loss questions produced very few usable results (6% - 50%). The Income Loss questions produced fewer LG responses (10% to 29%) than the Income Gain questions (19% to 44%).

The evidence seems to suggest that these responses are meaningful statements of preference. This is indicated by patterns in the LG responding and over/under trading e.g. LG responses fall as the level of income loss increases. If these responses were protest responses we might expect the prevalence to be constant across different income change levels. There is also consistency between questions, for example many respondents who would not trade any life years in the Income Loss questions would also not accept any amount of income in the Years Gain questions. Finally, there are a number of significant background characteristics that influence LG behaviour in an intuitive manner e.g. people in employment were less likely to give LG responses.

The above findings therefore do not suggest that the preference elicitation method is infeasible. Rather, it suggests that the questions are able to elicit meaningful responses, but the difficulty lies in incorporating some of these responses within the calculation method.

6.4.1.2 Reliability

The results of a simple test of logical consistency showed that the Income Loss questions produced the most consistent results. However, even for these questions only 55 out of 99 respondents gave a consistent ordering. A further 23 gave weakly consistent responses.

6.4.1.3 Validity

Both the Income Loss and Income Gain questions proved highly sensitive to scale. Referring back to the null hypotheses, the *first null hypothesis*, that the number of years traded in the Income Loss questions would not differ depending on the level of income loss, can be rejected. The *second null hypothesis*, that the number of years traded in the Income Gain questions would not differ depending on the level of income gain, can also be rejected. The Years Gain questions were also statistically significantly sensitive to scale, so the *third null hypothesis*, that the amount of income given up in the Years Gain questions would not differ depending on the number of years gained, can be rejected. The Years Loss questions were not statistically significantly sensitive to scale so the *fourth null hypothesis*, that the amount of income accepted would not differ depending on the number of years lost, cannot be rejected.

In line with expectations the Income Loss questions elicited significantly larger trades than the Income Gain questions, so the *fifth null hypothesis*, that there would be no differences, can be rejected. Also in line with expectations, amount of income required in the Years Loss questions, far exceeded the amount that was given up in the Years Gain questions, so the *sixth null hypothesis* can be rejected.

The MVQ estimates generated by the Income Loss, Income Gain and Years Gain are all correlated with ability to pay. This particular test is not of the validity of responses per se, but of the calculation method that is applied to responses, which differs from the application of this test of validity to Contingent Valuation. However, results of the test do suggest that the calculation method is effectively capturing the distribution of respondent income. The *seventh null hypothesis*, that the MVQ estimates would not differ by respondent income, can be rejected.

While the Income Loss questions have performed best in most of the tests, it is weakest in the simple test of convergent validity. Only 74% of responses in the 20% Income Loss questions were consistent with responses in the ranking exercise. The Years Loss questions had the most consistent responses (92%), while the 88% of the responses to the 20% Income Gain questions were consistent with responses to the ranking exercise.

6.4.1.4 MVQ Estimates

The key difficulty it seems is not in eliciting meaningful responses from participants, but in turning these responses into MVQ estimates. Using the individual calculation method for the Income Loss and Income Gain questions produces fairly consistent responses (although the two questions produce very different results) but this requires the exclusion of LG responses, of which there are a substantial amount. Using the aggregate calculation method allows incorporation of these respondents but represents a movement away from the theoretical roots of welfare economics.

The Years Gain questions are able to incorporate LG responses and produce consistent MVQ estimates. However, responses are severely restricted by the budget constraint, a finding which poses questions over the CV method once the benefit becomes sufficiently large. The Years Loss questions produced the most problematic responses due to the seemingly infinite preference for maintaining life years over increasing income.

6.4.2 Comparison with results from the Dutch online survey

This study was not designed specifically make comparison with the results from the Dutch online survey. The Dutch data was not originally intended for the purpose of deriving an MVQ, hence the study design was not optimal. In the UK study, rather than replicate the design of the Dutch study, a number of changes were made to improve the study design. For example, a visual aid was used as part of face to face interviews, and respondents were told they would be in the current state of health rather than perfect health. Furthermore, only the 20% Income Loss and 20% Income Gain questions have been asked in both studies. However, it is anecdotally interesting at this point to consider the results of the two studies in parallel.

There were far fewer LG responses in the UK study than in the Dutch online survey. There are a number of possible explanations for this. A likely cause of this difference is the difference in mode of administration. Respondents facing an interviewer perhaps gave more thought to the exercise and tried harder to give a meaningful response. The smallest unit of trade was also smaller in the UK study (1 week compared to 1 month), which may have encouraged people to trade. The prevalence of over-trading behaviour was similar between the two studies suggesting that the new visual aid made little difference.

The Income Loss and Income Gain questions were more sensitive to scale in the UK study. This is likely due to the fact that each respondent in the UK study valued multiple income change levels, which will have allowed them to value a given income change in relation to a previous income change levels they had valued in an attempt to be consistent. If LG responders are included the Dutch study elicits smaller mean trades than the UK study due to the high number of LG responses in the Dutch study. If LG responses are excluded the Dutch study elicits larger mean trades. There is no obvious explanation for this.

The MVQ results are typically higher in the UK study which is to be expected given that the UK study elicits lower mean trades. The results in the UK study show greater correlation with ability to pay, suggesting greater validity.

In the UK study respondents were told they would be in their own current health in each scenario, while in the Dutch study respondents were told to imagine they were in perfect health. There is no evidence to determine how this affected valuations. It is possible that additivity between health and income does not hold and respondents value income differently because their ability to enjoy it is hindered by their reduced health. Therefore, while respondents would trade less income because the years they are gaining are in poorer health, this may be cancelled out by the fact that they attach a lower value to income. However, it is overly-optimistic to think respondents thought processes are as complex as above and the improved results in the UK study suggest that having one less imaginary component made the exercise more manageable.

Overall, the results of the UK study show an improvement on those in the Dutch study, suggesting some of the modifications have been successful. However, a number of important weaknesses still exist.

6.4.3 Weaknesses of the Study

The sample size of the study is small (100). The aim of the study was only to test the methods, not to produce definitive MVQ estimates. However, the limited sample size does mean any conclusions are tentative. The findings need to be confirmed with a larger sample. In particular, it would be interesting to see if the proportions of LG responders is consistent if the questions are tested with a larger sample size.

The questions do not involve a quality of life trade off, hence it could be argued that the results should be interpreted as the monetary value of a life year rather than a QALY. The implications of this are discussed further in the following chapter.

Own health was measured on the VAS scale in order to avoid the need for two visual aids. This means that own health and the hypothetical valuations are not on the same measurement scale. Ideally future research should seek to elicit valuations of own health through TTO.

This study performed analysis using individual and household income because the questions simply stated 'own current income'. Most discussion has concentrated on individual income as it was felt this is most likely to be what respondents are considering. Future research needs to be explicit in the wording of the questions.

6.5 Conclusion

The results suggest that the Income Loss and Income Gain questions are feasible preference elicitation methods, that produce meaningful responses. However, a number of difficulties have been encountered in trying to translate these values into MVQ estimates. Respondents seem able to complete the questions and give responses that are sensitive to scope. While there are a number of LG responses and over-trades, the evidence suggests that these might be meaningful statements of preference rather than protest responses or driven by a lack of understanding. This suggests that the calculation method should take these responses into account, but the method proposed here has not been able to do so in a satisfactory manner. The individual calculation method is not able to incorporate LG responders, while over-traders can be incorporated through truncation at zero. An aggregate calculation method offers a way of including LG responders but this represents deviation from the principles of welfare economics, which makes the theoretical basis for the method weak.

Experimentation with two new questions has been predominantly unsuccessful. Responses to the Years Gain questions produce widely varying results driven by an apparent infinite preference for life years over income which the calculation method is not able to incorporate. While the properties of the responses to the Years Gain questions perform well, they must be viewed with caution due to the restrictions posed by the budget constraint. This finding is related to the well known issue of the influence of the budget constraint seen in CV studies. The Years Gain and Years Loss questions could potentially be improved if the size of the outcome being valued were considerably reduced (see next section).

The finding of an often infinite preference for length of life over income poses questions over the feasibility of calculating a value for a life extending QALY from an individual perspective. Future research may need to take a societal perspective to remove the respondents' self-interest from the valuation exercise (note: this would be at the cost of a movement away from welfare economics).

CHAPTER SEVEN

General Discussion and Conclusion

Cost-Effectiveness Analyses present results in terms of cost per QALY gained. However, there is no clear evidence on what monetary value members of the public place on a QALY. Currently, decision makers, such as NICE, operate seemingly arbitrary thresholds. An MVQ informed by general population values could allow decisions to better reflect public preferences.

This thesis has sought to explore a new method for deriving an MVQ, based on the TTO method of preference elicitation. Initially a literature review was conducted considering how the VOSL has been calculated and used in the public sector and how it has been used to estimate an MVQ. In the following chapter the TTO and CV methods were compared in terms of their feasibility, reliability and validity. This chapter highlighted that there are strengths and weaknesses of both methods, but suggested there could be merit in exploring the TTO method as an alternative to CV for estimation of an MVQ. The method was initially explored through a Dutch online survey, the results of which posed questions over the feasibility of the method. To explore whether these findings were the result of the online mode of administration or the questions themselves, a UK based interview study was conducted, which was preceded by a small scale pilot that reviewed the methods. The results showed improved performance in comparison to the Dutch survey, but a number of uncertainties and challenges remain.

This chapter considers the key findings, contributions to knowledge, policy implications, potential weaknesses, and areas for future research arising from the thesis. Finally, conclusions are drawn.

7.1 Key Findings

This thesis initially tested two new questions to estimate an MVQ through a Dutch online survey. These questions asked respondents to trade off length of life to either maintain their current income or increase their income. These questions were further tested through a small scale pilot study and a UK based interview study. In addition two further variants of the questions were also tested, in which respondents either gave up income to increase their life expectancy or accepted an increase in income to compensate for a fall in life expectancy.

7.1.1 Feasibility

In the UK interview study only one interview had to be terminated because the respondent did not understand the task, and if LG responders are treated as completers this results in a completion rate of 99%.

Responses in which respondents stated an infinite preference for length of life over income (LG responses) were the greatest problem with the methods. There was a very high prevalence of LG responses in the Income Loss and Income Gain questions in the Dutch study. Even after excluding the 'extreme non-traders' (did not trade in all 14 TTO exercises) 56% of Income Loss responses and 64% of Income Gain questions were LG responses. There were considerably less LG responses in these two questions in the UK study. Only 17% of Income Loss responses and 30% of the Income Gain responses were LG responses. The Years Loss questions were more problematic, with 71% prevalence of LG responses. A total of 15% of the Years Gain responses were LG.

The prevalence of LG responses in the TTO based questions was higher than the rates typically seen in the CV studies reviewed in Chapter 3. However, the evidence suggests that these responses may be meaningful statements of preference rather than protest responses. In the UK study, for a given question type (e.g. Income Loss), the prevalence of LG responses tended to increase as the level of income change decreased. If these were protest responses one would expect the prevalence to be constant across income change levels. Also, many respondents who would not trade any years in the Income Loss questions would not accept any amount of income in the Years Loss questions.

These findings suggest that respondents have a very strong preference for length of life over income. Referring to the Indifference Curve Analysis presented in Chapter 4 (Figure 4.2), respondents' indifference curves are vertical lines (or very steep lines) at the particular point they are asked to make a trade i.e. they have lexicographic preferences. This suggests that the budget constraint may be limiting respondents' willingness to trade i.e. the size of the outcome being valued is too large (Years Gain and Years Loss) or too small (Income Gain and Income Loss). In the Years Gain questions respondents value the additional years of life more than the total amount of their income. Likewise, in the Years Loss questions there is no amount of money respondents would accept to reduce their life expectancy by such a large amount. In the Income Loss and Income Gain questions respondents would not accept a fall in life expectancy to increase their income. The limiting effect of the budget constraint has been acknowledged in both the CV studies eliciting an MVQ in chapter 2, and the broader CV literature reviewed in Chapter 3. CV studies try to avoid this problem by reducing the size of outcome e.g. by valuing small risk reductions. Ways in which the TTO based method could potentially be adapted to address this problem will be discussed in section 7.4 below.

A total of 17% of Income Loss responses and 19% of Income Gain questions were over-trades in the Dutch study. The percentage in the UK study was similar for the Income Loss questions with a total of 16% of responses being over-trades, but was lower for the Income Gain questions with 9% of responses being over-trades. A total of 7% of the Years Loss responses were under-trades, while 19% of the Years Gain responses were under-trades.

There are a number of possible explanations for over/under trading behaviour: respondents do not understand the exercises; respondents understand the exercises but are not able to determine the point at which lifetime income becomes lower in the alternative scenario; the assumption of additive separability between health and income does not hold. It is difficult, based on the results of the studies, to determine which explanation is most likely/common.

Feedback in the UK study suggested respondents understood the questions well, but did not try to calculate total lifetime income in each of the scenarios. Interestingly, the regression analysis in Chapter 6 showed that respondents who did try to calculate total lifetime income traded significantly more years in both the Income Loss and Income Gain questions. Calculating total lifetime income is potentially complex depending on the particular combination of life years and percentage of income. Ways of improving this will be addressed in the 'future research' section.

7.1.2 Reliability

Reliability has only been tested through a consistency test through comparisons with a ranking exercise in the UK study. Only 23 out of 99 responses were inconsistent. Reliability needs to be tested further in any further research (see section 7.4 below).

7.1.3 Validity

The results suggest that sensitivity to scale is an area in which the TTO method may improve on the CV method. In the Income Loss questions respondents faced with a larger income loss traded, on average, a greater number of years. This held in both the Dutch study and the UK study, although the UK study displayed greater sensitivity to scale. The Income Gain questions were insensitive to scale in the Dutch online survey. However, they were sensitive to scale in the UK study. The Years Gain and Years Loss questions were both sensitive to scale. In contrast CV studies deriving an MVQ (e.g. Prades *et al.* 2009, Donaldson *et al.* 2011) and some CV studies in the healthcare field have found the CV method to be insensitive to the size of the health gain being valued (e.g. Olsen *et al.* 2004, although the evidence reviewed in Chapter 3 is mixed).

The MVQ estimates in the UK study are well correlated with ability to pay, but this is not the case in the Dutch study. The evidence reviewed earlier showed that the CV method performs well against this test of validity. However, the tests are applied to the two methods differently. In the CV method correlation with ability pay should reveal itself directly in people's responses. In the TTO method application of the calculation method to the responses should reveal a correlation with ability to pay due to the role of respondent income in the calculation method.

7.1.4 Comparing the four questions

In the Years Loss questions the size of the outcome being valued was too large, meaning that a large number of respondents were not prepared to accept any amount of income as compensation (preventing the calculation of an MVQ). In the Years Gain questions the size of the outcome was again too large, meaning that responses were strongly restricted by the budget constraint. This does not necessarily mean that the questions are infeasible or invalid, but that any further testing of the methods would need to considerably reduce the size of the outcome being valued.

Comparing the Income Loss and Income Gain questions, the Income Loss elicited the lowest proportion of LG responses and is most sensitive to scope.

7.1.5 MVQ results

In this section I focus on the results from the UK study as this appears to have produced the more robust results. Results have been presented through both the individual and aggregate approaches. Given that early evidence suggests that LG responses may have been true statements of preference rather than protest responses, the calculation method should ideally be able to incorporate these responses. This would require the use of the aggregate approach. However, this does not fully account for individual preferences. On the other hand, the individual approach requires the exclusion of LG responses, requires the truncation of the values of the over/under-traders, and is very vulnerable to the smallest unit of trade. The issues over which calculation method is appropriate have been discussed in the context of studies eliciting a health state valuation and a CV derived WTP valued, with the aim of combining them to produce an MVQ (Gyrd-Hansen and Kjaer, 2012). No firm conclusions have been drawn. The MVQ results presented here, like the CV based MVQ results, are higher when calculated through the individual approach, due to the effect of marginal responses.

The Income Loss questions in the UK study produced MVQ estimates, calculated through the individual approach, ranging from £70K to £91K. When calculated through the aggregate approach the estimates ranged from £20K to £26K. The individual estimates are considerably higher than the existing NICE threshold, while the aggregate estimates lie within the threshold. This illustrates the importance of the decision over whether to use the individual or aggregate calculation method. The chained approach in the EuroVaQ project produces values ranging from £8,200 to £37,400 for the UK, depending on the particular question and health state valued. Further comparison with the existing MVQ estimates derived from the public (rather than the existing VPF) presented in Table 2.2 in Chapter 2 show that the individual estimates from the Income Loss questions are high. Only the study by Prades *et al.* (2009) and the SVQ project (if outliers are included, Baker *et al.* 2010) produce higher estimates (note, both of these also use the individual calculation method and are susceptible to the influence of small trades). This is most likely driven by the influence of small trades in the Income Loss questions.

In the EuroVaQ project (Donaldson, 2011), the largest study eliciting an MVQ to date, a direct approach is used as well as a chained approach. The two questions most similar to the TTO based questions asked respondents their WTP for additional months of life at the end of life, or their WTP for additional months now. Respondents are told they will be in their own current health for these months, and a number of months is chosen that corresponds to one QALY. Calculation is through the individual approach. These questions are very similar to the Years Gain questions used in the UK based study, although the outcome being valued in the Years Gain questions is larger. Even the smallest increase in years (2) is considerably larger than the 1 QALY gain valued in the EuroVaQ project (exact size depends on valuation of own health). The EuroVaQ questions asking for WTP for additional months of life at the end of life produce estimates for the UK of £6,000. The questions asking for WTP for additional months of life now produce estimates of £26,000. In comparison, the 2 Years Gain questions produce estimates of £62,450 using the individual calculation method. This is somewhat surprising given that the EuroVaQ authors cite the influence of the budget constraint in their questions. One would expect the influence of the budget constraint to be even larger in the Years Gain questions since the outcome being valued is larger, which would lead to lower values. One possible explanation for this difference is that in the EuroVaQ questions only between 53% and 79% agreed to pay for the health gain (depending on the particular question). The majority of these indicated at least one reason that indicated they valued the health gain as zero. In contrast, in the 2 Years Gain questions no respondents refused to give up any income. Shiroiwa *et al.* (2010) also asked respondents their WTP to avoid immediate death and have one additional year of life in full health. They produced a value of £22,000 for the UK which is again lower than the individual estimates produced by the Years Gain questions.

Given that the TTO based methods do not include a quality of life trade off the value of a life year derived from the existing VOSL is perhaps a more relevant comparator. Mason *et al.* (2009) using the existing UK VOSL (and the simplistic calculation method) produced a value of a life year of £73,000 which is very similar to the values generated by the Income Loss questions, using the individual calculation method.

The Income Gain questions produce higher values than the Income Loss questions. This is consistent with diminishing marginal utility of income. Incremental increases in income decrease in value, so fewer life years are given up, which generates a higher MVQ estimate. Considering the MVQ results from the Years Gain and Years Loss questions allows comparison with the observed discrepancy between WTP and WTA in CV studies as discussed in Chapter 3. The Years Loss questions (equivalent to WTA) produce considerably higher estimates than the Years Gain questions (equivalent to WTP). Thus, the results here are consistent with the finding in the CV literature that WTA produces higher results than WTP. The difference is particularly pronounced in the Years Gain and Years Loss questions due to the strong effect of diminishing marginal utility of both income and life years, and also by the strong budget constraint in the Years Gain questions.

7.2 Key weaknesses of the MVQ studies

As mentioned above the LG responses and over/under trades were problematic. These responses in themselves are not a weakness *per se*, rather it is a shortcoming of the calculation method that these responses cannot be incorporated effectively. Furthermore, a weakness of the study design was that it was not possible to strongly determine whether (or what proportion of) these responses were meaningful statements of preference, protest responses or due to a lack of understanding of the tasks.

The initial aim of the Dutch online survey was to address the inclusion of productivity costs in TTO valuations of EQ-5D states. The Income Loss and Income Gain questions were then used opportunistically to derive an MVQ. For this reason, a larger scale online survey preceded the pilot and subsequent interview study. Had study design been aimed at addressing the issue of an MVQ from the outset the pilot could have been conducted first, followed by the interview study. At this stage, a decision could have been made as to whether or not the questions were likely to be feasible through an online mode of administration.

The sample size of the UK based study is small as it was limited by research budget constraints. The aim was to explore the questions, not to produce definitive MVQs. However, one must be careful when attempting to draw conclusions on factors such as the number of LG responses, and sensitivity to scale.

The sampling was not particularly rigorous. The interviewers drew upon a pool of known respondents, many of whom had completed TTO exercises before. However, given the innovative nature of the questions it is perhaps unlikely that responses will have been contaminated by previous experience of TTO exercises. It is more likely that the £25 incentive offered to respondents may have influenced responses. This may provide an explanation for the reduced prevalence of LG responses in the UK based study when compared to the Dutch online survey (in which respondents could only receive a donation to a charity of their choice). Respondents when faced with an interviewer, and in the knowledge that they are being paid well for their time, may feel more obliged to "play the game" by trading at least some time/income.

In both the Dutch online survey and the UK study it was not explicitly mentioned whether respondents should consider individual income or household income. Analysis has been performed using both, but discussion has focused on results based on individual income since it is believed 'own current income' is more closely associated with individual income. Future work should be more explicit in the wording of the questions.

It could be seen as a weakness that the methods do not involve a quality of life trade off. The implications of this are discussed in the next section.

7.3 Contribution to knowledge and Implications for Policy

This thesis has sought to test an exploratory method for deriving an MVQ as an alternative to CV. Results of the UK interview study suggest the questions are acceptable to respondents and are potentially more sensitive to scale than CV studies. The results show that the Income Loss questions are the most promising, but a number of problems still exist. Most notably, the issues of LG responding and over-trading need to be addressed in any further research (see below section 7.4).

Results of the method need to be viewed in light of the fact that there is not quality of life trade off. This is not the first study to elicit the value of a QALY exclusively through life extensions, and if one adheres to the notions of the QALY model that all QALYs are equal regardless of whether they are derived from quality of life improvements or life extensions, the results of the TTO based methods should be able to be applied universally. However, existing studies eliciting an MVQ have shown that the value differs depending on the particular outcome being valued. The implication is that the results of the TTO based method should only be applied to life extending treatments. Alternatively, given that there seems to be a range of MVQ estimates and life extensions are valued more highly then quality of life improvements, the results of the method could inform the upper limit of the range. Comparison with existing values shows that the results of the Income Loss questions are most closely aligned with the value of a life year derived from the public sector VOSL. In a scenario where a value of a life year is to be applied in a healthcare context it may be preferable to use a value that is informed by public preferences on life extensions, rather than the WTP to reduce the risk of death caused by means unrelated to health care.

It seems unlikely that a Cost Effectiveness threshold would be informed directly from an MVQ informed by public preferences (regardless of the method), predominantly due to fiscal restrictions. In reality it seems that information on the MVQ is more likely to be used to give an indication to government as to whether the budget is set to appropriately reflect public opinion, rather than to be applied rigorously by decision making bodies as a cost effectiveness threshold. Not only does the potential for creating a 'one size fits all' value seem limited, but such an approach could have serious implications for the budget that would have consequences for other areas of the public sector. This does not mean that researchers should be deterred from further exploring the methods for deriving an MVQ as existing work still gives no clear indication of whether current thresholds accurately reflect public opinion.

7.4 Areas for Future Research

Put simply, LG responses in the Income Loss questions arise because what is being given up (the number of life years) is deemed to be more valuable than what is being gained (the amount of income). There are two ways that this problem can be addressed: either the volume of what is being gained can be increased, or the volume of what is being given up can be decreased. The total value of what was being given up was lower in the UK study than the Dutch study because current health was used instead of perfect health (unless respondents consider themselves to be in perfect health). This might offer one explanation for why prevalence of LG responding decreased in the UK study. A way in which both the volume of what is being given up could be decreased and the volume of what is being gained could be increased would be by increasing the time horizon. For example, if the time horizon was increased to 30 years the total value of years given up at the end of life would decrease due to both diminishing marginal utility of life years and positive rate of time preference. At the same time, the total value of what is being gained would increase because an increase in annual income would cause a larger increase in lifetime income. One with such an approach would be the influence of time preference, since the loss in years occurs at the end of life, but the increase in income occurs on an annual basis. Adjustment for time preference might be required which would complicate both the task and the calculation method.

An alternative way in which the total value of what is being given up could be decreased is to specify that respondents are in a given sub-optimal health state for the given time horizon. However, this might complicate the task by introducing a further hypothetical element.

The new visual aid used in the UK based study seems to have had little effect on the prevalence of over/under-trading behaviour. This may be because income in each of the two scenarios is presented as annual income, making it difficult to determine lifetime incomes, and hence difficult to determine the point at which lifetime income in a given scenario becomes lower than in the other scenario. A computer visual aid that is able to calculate lifetime income in each of these scenarios may be able to overcome this problem. However, over/under-trading behaviour may not be caused by an inability to calculate lifetime income in each of the two scenarios. Respondents may simply prefer to live for a shorter period of time with a higher annual income, even though lifetime income is lower (i.e. the relationship between health and income may not be additively separable).

The smallest unit of trade is an important decision in any future research. Unfortunately the results obtained seem to be sensitive to 'framing effects'; i.e. influenced by the options given. Using a smallest unit of trade of one week in the UK study generated some very large estimates. There is a trade off between wanting a very small unit of trade to avoid LG responses, but not wanting it to be too small to avoid generating very large estimates that heavily influence the mean. The choice of the smallest unit of trade may influence whether the individual or aggregate calculation approach is most appropriate. The aggregate approach dilutes the effect of very small trades. In the individual approach one must either allow these very high values to disproportionately affect the means, employ an arbitrary truncation or exclude these observations.

Reliability of the method has not yet been adequately addressed. Ideally this would be done through test-retest reliability, with the focus on stability of means as this is most relevant from a policy perspective.

In conclusion, this thesis recommends that future research should:

- Focus on the Income Loss questions;
- Use the interview method of administration;
- Consider experimenting with a computer visual aid that can present respondents with lifetime income in each scenario;
- Word the questions to explicitly instruct respondents to consider individual, not household, income;
- Directly test the effect of differing the smallest unit of trade;
- Experiment with extending the time horizon; and
- Test reliability through a test-retest design.

7.5 Conclusion

Experimentation with four different types of TTO based questions to elicit an MVQ has identified the Income Loss questions as a potential new method for the estimation of an MVQ. The questions seem to be acceptable to respondents and sensitive to scale. A number of weaknesses still exist, including LG responding and over-trading, and a number of proposals for how this could be addressed in further research have been made. Through further development the method could potentially be used to contribute to the growing literature informing a likely range of MVQ estimates to aid understanding of how existing CE thresholds align with public preferences.

References

Abdellaoui, M, Barrios, C, Wakker, PP. 2000. Did vNM resurrect cardinal utility after all? Theoretical and empirical arguments based on non-expected utility. Working Paper, ENS, Cachan.

Abelson, P. 2003. The value of life and health for public policy. *The Economic Record*, **79**: S2-S13

Acton JP. Evaluating public programs to save lives: the case of heart attacks. Santa Monica: Rand. Report R-950-RC, 1973.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory,* ed. B. Petrov and F. Csaki, 267-281. Budapest: Akademiai Kiado.

Aldy, J.E., and W.K. Viscusi. 2007. Age Differences in the Value of Statistical Life: Revealed Preference Evidence. *Resources for the Future, Discussion Papers.*

Aldy, J.E., and W.K. Viscusi 2008. "Adjusting the Value of a Statistical Life for Age and Cohort Effects." *Review of Economics and Statistics* 90(3): 573–81.

Anderson G, Black C, Dunn E, Alonso J, Christian-Norrega J, Folmer-Anderson T, Bernth-Peterson, P. 1997 Willingness to pay to shorten waiting time for cataract surgery. *Health Affairs,*16:181–91.

Appel LJ, Steinberg EP, Powe NR, Anderson GF, Dwyer SA, Faden RR. 1990; Risk reduction from low osmolarity contrast media: what do patients think it is worth? *Medical Care*, **28**: 324–37.

Appleby, J., Devlin, N., Parkin, D., Buxton, M., Chalkidou, K. 2009. Searching for cost effectiveness thresholds in the NHS. *Health Policy,* **91** (3): 239-245

Arabsheibani, G.R., Marin, A., 2000. Stability of estimates of the compensation for danger. Journal of Risk and Uncertainty 10 (3), 247–269.

Arnesen T, Norheim OF. Quantifying quality-of-life for economic analysis: time-out for Time trade-off. J Med Eth Med Hum 2003; 29: 81–86.

Arnesen TM, Trommald M. 2004. Roughly right or precisely wrong? Systematic review of quality-of-life weights elicited with the time trade-off method. *J Health Serv Res Pol,* 9(1): 43–50.

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

Arnould, R.J. and L.M. Nichols. 1983. "Wage-Risk Premiums and Workers' Compensation: A Refinement of Estimates of Compensating Wage Differential," *Journal of Political Economy* 91(2), 332–340.

Arrow, Kenneth J., (1951). *Social Choice and Individual Values*. Wiley, New York.

Arrow, K., Solow, R., Portney, P.R., Leamer, E.E., Radner, R., Schuman, H. *Report to the NOAA panel on contingent valuation*. Mimeo; 1993.

Asenso-Okyere WK, Osei-Akoto I, Anum A, Appiah EN. Willingness to pay for health insurance in a developing country. A pilot study of the informal sector of Ghana using contingent valuation. Health Policy 1997; 42: 223–37.

Ashby, J., O'Hanlon, M., Buxton, M.J. 1994. The Time trade-off technique: how do the valuations of breast cancer patients compare to those of other groups? *Quality of Life research,* **3:** 257-65.

Ashenfelter, O. 2005. Measuring the Value of a Statistical Life: Problems and Prospects. Working Papers (Princeton University. Industrial Relations Section).

Atkinson SE, Halvorsen R. 1990. The Valuation of Risks to Life: Evidence from the Market for Automobiles. *Review of Economics and Statistics,* **72** (1): 133-36

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

Attema, A.E., Brouwer, W.B.F. 2011. The way that you do it? An elaborate test of procedural invariance of TTO, using a choice based design. *European Journal of Health Economics,* Epub ahead print.

Badia, X., Roset, M., Herdman, M. 2001. Feasibility and validity of the VAS and TTO for eliciting general population values for temporary health states: A comparative study. *Health Services and Outcomes research Methodology,* **2:** 51-65.

Badia, X., Roset, M., Herdman, M. 1999. Inconsistent responses in three preference elicitation methods for health states. *Social science and medicine,* **49:** 943-950.

Baker R, Bateman I, Donaldson C, Jones-Lee M, Lancsar E, Loomes G, *et al.* 2010 Weighting and valuing quality-adjusted life-years using stated preference methods: preliminary results from the Social Value of a QALY Project. *Health Technol Assess;* **14** (27).

Baker, R., Chilton, S., Jones-Lee, M., Metcalf, H. 2008. Valuing lives equally: Defensible premise or unwarranted compromise? *Journal of Risk and Uncertainty,* **36:** 125-138

Bala MV, Wood LL, Zarkin GA, Norton EC, Gafni A, O'Brien BJ.1999. Are health states "timeless"? The case of the standard gamble method. *Journal of Clinical Epidemiology*, **52:** 1047-1053.

Baron, J. 1997. "Confusion of Relative and Absolute Risk in Valuation," *Journal of Risk and Uncertainty* 14, 301–309.

Bateman, I.J., Carson, R.T., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones, A., Loomes, G., Mourato, S., Ozdemiroglu, E., Pearce, D.W., Sugden, R., Swanson, J. Economic valuation with stated preference techniques: a manual. Cheltenham: Edward Elgar; 2002.

Bellavance F, Dionne G, Lebeau M. 2009. The Value of a Statistical Life: A Meta-analysis with a Mixed Effects Regression Model. *Journal of Health Economics*, 28 (2): 444-64

Berger, M.L., Murray, J.F. 2001. Alternative valuations of work loss and productivity. *Journal of Occupational and Environmental Medicine,* **43** (1): 18-24.

Berwick DM, Weinstein MC. What do patients value? 1985. Willingness to pay for ultrasound in normal pregnancy. *Medical Care*, **7**: 881–93.

Bhatia, M. 2005. From evidence to calibration for starting point bias: willingness to pay for treated mosquito nets in Gujarat, India. *Applied Economics,* **37:** 1-7.

Bhatia, M. and Fox-Rushby, J. 2002. Willingness to pay for treated mosquito nets in Surat, India: the design and descriptive analysis of a household survey, Health Policy and Planning, 17, 402–11.

Birch, S., Gafni, A. 2002. On being NICE in the UK: guidelines for technology appraisal for the NHS in England and Wales. *Health Economics,* **11** (3): 185-191.

Blomquist GC. 2004. Self-Protection and Averting Behavior, Values of Statistical Lives, and Benefit Cost Analysis of Environmental Policy. Review of Economics of the Household, **2** (1): 89-110.

Blumenschein, K., Johannesson, M. 1998. Relationships between quality of life instruments, health state utilities and willingness to pay in patients with asthma. *Annals of Allergy, Asthma and Immunology,* **80:** 189-194.

Bowland, B.J., Beghin, J.C. 2001. Robust estimates of value of a statistical life for developing economies. *Journal of Policy Modeling,* **23:** 385-396.

Bishop, R.C., Heberlein, T.A. 1979. Measuring values of extra-market goods: are indirect measures biased? *American Journal of Agricultural Economics,* **61:** 926-930

Black, W.C. 1990. The cost-effectiveness plane: a graphic representation of cost-effectiveness. *Medical Decision Making,* **10:** 212-215.

Black, D.A., J. Galdo, and L. Liu. 2003. "How Robust are Hedonic Wage Estimates of the Price of Risk?" Prepared for the U.S. Environmental Protection Agency.

Bland, J.M., Altman, D.G. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet,* **1:** 307-310.

Bland, J.M., Altman, D.G. 1996. Measurement error and correlation coefficients. *BMJ,* **313:** 41-42.

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., Filko, M., 2008. New tests of QALYs when health varies over time. *Journal of Health Economics,* **27:** 1237–1249.

Bleichrodt, H., Filko, M. 2010. A reply to Gandjour and Gafni. *Journal of Health Economics,* **29:** 329-331.

Bleichrodt, H., Johannesson, M. 1997. The validity of QALYs: an experimental test of constant proportional time trade-off and utility independence. *Medical Decision Making,* **17** (1): 21-32.

Bleichrodt H, Pinto JL. 2000. A parameter-free elicitation of the probability weighting function in medical decision analysis. Management Sciences; **46**: 1485–1496.

Bleichrodt, H., Pinto, J. L., & Abellan-Perpinan, J. M. 2003. A consistency test of the time trade-off. *Journal of Health Economics,* **22**: (6), 1037–1052.

Blomqvist, G.C., Dickie, M., O'Conor, R.M. 2011. Willingness to pay for improving fatality risks and asthma symptoms: values for children and adults of all ages. *Resource and Energy Economics,* **33** (2): 410-425.

Blumenschein, K., Johannesson, M., Yokoyama, K. K., & Freeman, P. R. 2001. Hypothetical versus real willingness to pay in the health care sector: Results from a field experiment. Journal of Health Economics, **20**(3), 441–458.

Bobinac, A., van Exel, N.J.A., Rutten, F.F.H., Brouwer, W.B.F. 2010. Willingness to Pay for a Quality Adjusted Life Year: The Individual Perspective. *Value in Health,* **13** (8): 1046-1055.

Broadway, R.W., Bruce, N. *Welfare Economics.* Blackwell, Oxford, 1984.

Brooks, R.G., Jendteg, S., Lindgren, B., Persson, U., Bjork, S. 1991. EuroQol: health related quality of life measurement. Results of the Swedish questionnaire exercise. *Health Policy,* **18:** 37-48.

Brooks, R. with the EuroQoL Group. 1996. EuroQol: the current state of play. *Health Policy,* **37:** 53-72.

Brouwer, W.F.B., van Exel, J., Baker, R., Donaldson, C. 2008. The New Myth: The Social Value of the QALY. *Pharmacoeconomics,* **26** (1): 1-4

Brown, C. 1980. Equalizing Differences in the Labor Market. *Quarterly Journal of Economics,* **94** (1): 113-134.

Buckingham JK, Birdsall J,Douglas JG. 1996. Comparing three versions of the time tradeoff: time for a change? *Med Decis Making.*; **16** (4): 335–47.

Buckingham, K., Devlin, N. 2006. A theoretical framework for TTO valuations of health. *Health Economics,* **15:** 1149-54.

Broome J. 1993. QALYS. *Journal of Public Economics*, 50 (**2**): 149–67.

Byrne, M.M., O'Malley, K., Suarez-Almazor, M.E. 2005. Willingness to pay per quality-adjusted life year in a study of knee osteoarthritis. *Medical Decision Making,* **25** (6): 655-666.

Cairns JA, van der Pol MM. Saving future lives: a comparison of three discounting models. *Health Economics Journal* 1997; **6**: 341–350.

Camerer, CF., Ho, T-H. 1994. Non-linear weighting of probabilities and violations of the betweenness axiom. Journal of Risk and Uncertainty, 8: 167–196.

Carson, R.T., Flores, N., Hanemann, W. 1998. Sequencing and valuing public goods. *Journal of Environmental Economics and Management,* **36:** 314-323

Carson, R.T., Hanemann, W.M., Mithcell, R.C. 1986. Determining the Demand for Public Goods by Simulating Referendums at Different tax Prices. Department of Economics Working Paper, University of California, San Diego.

Carthy, T., Chilton, S., Covey, J., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N., Spencer, A. 1999. On the contingent valuation of safety and the safety of contingent valuation: part 2 – the CV/SG "chained" approach. *Journal of Risk and Uncertainty,* **17** (3): 187-213.

Cameron TA. 2010. Euthanizing the Value of a Statistical Life. Review of Environmental Economics and Policy: **4** (2): 161-78.

Chestnut, K.L., Lambert, W., Rowe, R. 1996. Measuring heart patient' willingness to pay for changes in angina symptoms. *Medical Decision Making,* **16(1):** 65-77

Chestnut, L. G., R.D. Rowe, J. K. Lazo, andW. S. Breffle. 2003. *Economic Valuation of Mortality Risk Reduction: Stated Preference Approach.* Report prepared for Cornell University and the USEPA. Cooperative Agreement CR824393-01-0. Stratus Consulting, Inc., Boulder, CO.

Chilton, S., Covey, J., Jones-Lee, M., Loomes, G., Metcalf, H. 2004. *Valuation of health benefits associated with reductions in air pollution.* London: DEFRA.

Chilton, S., Covey, J., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N., Spencer, A. 2002. Public perceptions of risk and preference based values of safety. *Journal of Risk and Uncertainty,* **25:** 211-232.

Churchill, D.N., Morgan, J., Torrance, G.W. 1984. Quality of life in end stage renal disease. *Peritoneal Dialysis Bulletin,* **4:** 20-23.

Churchill, D.N., Torrance, G.W., Taylor, D.W., Barnes, C.C., Ludwin, D., Shimizu, A., Smith, K.M. 1987. Measurement of quality of life in end stage renal disease: The time trade-off approach. *Clinical and Investigative Medicine,* **10:** 14-20.

Cho, M. N., Lertmaharit, S., Kamol-Ratanakul, P., & Saul, A. J. 2000. Ex post and ex ante willingness to pay(WTP) for the ICT Malaria Pf/Pv test kit in Myanmar. *Southeast Asian Journal of Tropical Medicine and Public Health,* **31**, 104–111.

Coley CM, Li Y-H, Medsger AR, Marrie TJ, Fine MJ, Kapoor WN, Lave JR, Detsky AS, Weinstein MC, Singer DE. 1996. Preferences for home vs. hospital care among low-risk patients with community-acquired pneumonia. *Archives of Internal Medicine,* 156: 1565–71.

Cook J, Richardson J, Street A. 1994. Acost-utility analysis of treatment options for gallstone disease—methodological issues and results. *Health Economics Journal,* **3** (3): 157–68.

Cookson, R. 2003. Willingness to pat methods in health care: a sceptical view. *Health Economics,* **12:** 891-894.

Corso, P.S., J.K. Hammitt, and J.D. Graham. 2001. "Valuing Mortality-Risk Reduction: Using Visual Aids to Improve the Validity of Contingent Valuation." *Journal of Risk and Uncertainty* **23**(2): 165–84.

Court, C. 1995. Survey shows widespread rationing in NHS. *BMJ,* **311:** 1453-4

Cousineau, J.-M., Lacroix, R., Girard, A.M., 1992. Occupational hazard and wage compensating differentials. *The Review of Economics and Statistics* **74,**
166–169.

Covey, J., Loomes, G., Bateman, I. 2007. Valuing Risk Reductions: Testing for Range Biases in Payment Card and Random Card Sorting Methods. *Journal of Environmental Planning and Management,* **50** (4): 467-482.

Covey, J., Loomes, G., Swann, M. 2002. Taking a hint: some anchors are more informative than others. Paper presented at the Society for Judgement and Decision Making Annual Meeting, Kansas City, MO, 24-25 November.

Cropper, M., Hammitt, J.K., Robinson, L.A. 2011. Valuing Mortality Risk Reductions: Progress and Challenges. NBER Working Paper No. 16971.

Culyer, A., McCabe, C., Briggs, A., Claxtonm K., Buxton, M., Akehurst, R., Sculpher, M., Brazier, J. 2007. Searching for a threshold, not setting one: the role of the National Institute for Health and Clinical Excellence. *Journal of Health Services Research and Policy,* **12** (1): 56-58.

Cunningham, S.J., Hunt, N.P. 2000. Relationship between utility values and willingness to pay in patients undergoing orthognathic treatment. *Community Dental Health,* **17:** 92-96.

CVZ. 2006. "Dutch guidelines for pharmacoeconomic research ("Dutch board of insurers")", *College Voor Zorgverzekeringen*, available at: www.ispor.org/PEguidelines/source/PE_guidelines_english_Netherlands.pdf.

Dardis, R. 1980. The Value of Life: New Evidence from the Market Place. *American Economic Review,* **70:** 1077-82.

De Blaeij A, Florax R. 2003. The Value of Statistical Life in Road Safety: A Meta-Analysis. *Accident Analysis and Prevention*, 35(6): 973-986

Dekker, T., Brouwer, R., Hofkes, M., Moeltner, K. 2011. The Effect of Risk Context on the Value of a Statistical Life: A Bayesian Meta-Model. *Environmental and Resource Economics.* **49:** 597-624.

Delquie P. 1993. Inconsistent trade-offs between attributes: new evidence in preference assessment biases. *Management Sciences*, **39**: 1382–1395.

Delquie P. 1997. 'Bi-Matching': a new preference assessment method to reduce compatibility effects. *Management Sciences*, **43**: 640–658.

Department of Health. 1999. *Economic Appraisal of the Health Effects of Air Pollution.* London: The Stationary Office.

Department for Transport. 2007. *Highways Economics Note No. 1: 2005,* Department for Transport, London.

Department for Transport. 2011. Updating the Value of a Prevented Fatality (VPF) and Values of Prevented Injuries (VPI). Department for Transport, London.

Desaigues, B., Rabl, A. Reference values for human life: an economic analysis of a contingent valuation in France. In: Schwab Christe, N.G., Soguel, N.C. (eds). Contingent Valuation, Transport Safety and the Value of Life. Boston, MA: Kluwer Academic Publishers, 1995; 85-112.

Detsky, A.S., McLaughlin, J.R., Abrams, H.B., L'Abbe, K.A., Whitwell, J., Bombardier, C., Jeejeebhoy, K.N. 1986. Quality of patients on long term total parenteral nutrition at home. *Journal of General Internal Medicine,* **1** (1): 26-33.

Diener, A, O'Brien, B, Gafni, A. 1998. Health care contingent valuation studies: a review and classification of the literature. *Health Economics,* **7**: 313–26.

Dillingham, A.E., Smith, R.S., 1984. Union effects on the valuation of fatal risk. In: Proceedings of the Industrial Relations Research Association 36th Annual Meeting, San Francisco.

Dionne G, Lanoie P. 2004. Public choice about the value of a statistical life for cost-benefit analyses - The case of road safety. Journal of Transport Economics and Policy, **38**: 247-74.

Dolan, P. 1997. Modelling Valuations for EuroQol Health States. *Medicare,* **35**(11): 127-156

Dolan, P. 1998. Valuing Health-Related quality of life: Issues and Controversies. *Pharmacoeconomics,* **15** (2): 119-127.

Dolan, P., Edlin, R. 2002. Is it really possible to build a bridge between cost-benefit analysis and cost-effectiveness analysis? *Journal of Health Economics,* **21:** 827-843

Dolan P, Gudex C. 1995. Time preference, duration and health state valuations. *Journal of Health Economics* **4**: 289–299

Dolan, P., Gudex, C., Kind, P., Willams, A. 1996a. Valuing Health States: A comparison of methods. *Journal of Health Economics,* **15:** 209-231.

Dolan, P., Gudex, C., Kind, P., Willams, A. 1996b. The Time Trade-Off method: results from a general population study. Journal of *Health economics,* **5:** 141-154.

Dolan P, Jones-Lee M .1997. The time trade-off: a note on the effect of life time real location of consumption and discounting. Journal of Health Economics 16: 731–739

Dolan, P., & Stalmeier, P. F. M. 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.

Donaldson, C., Baker, R., Bateman, I *et al.* Weighting and valuing quality adjusted life years: preliminary results from the Social Value of a QALY project. In: *Report for the Methodology Programme.* NIHR, Birmingham, UK, 2008.

Donaldson, C., Baker, R., Mason, H. and the EuroVaQ Team. 2011. European Value of a Quality Adjusted Life Year. Available online: http://research.ncl.ac.uk/eurovaq.

Donaldson, C., Farrar, S., Mapp, T., Walker, A., Macphee, S. 1997. Assessing Community Values in Health Care: Is the Willingness to Pay Method Feasible? *Healthcare analysis,* **5**(1): 7-29.

Donaldson, C., Hundley V., Mapp, T. 1998. Willingness to Pay: A Method for Measuring Preferences for Maternity Care? *Birth,* **25** (1): 32-39.

Donaldson C, Shackley P, Abdalla M, Miedzybrodzka Z. 1995. Willingness to pay for antenatal carrier screening for cystic fibrosis. *Journal of Health Economics*; **4**: 439–52.

Dong, H., Kouyate, B., Cairns, J., Sauerborn, R. 2003. A comparison of the reliability of the take-it-or-leave-it and the bidding game approaches to estimating willingness-to-pay in a rural population in West Africa. *Social Science and Medicine,* **56:** 2181-2189.

Dreyfus, M.K., Viscusi, W.K. 1995. Rates of Time Preference and Consumer Valuation in Automobile Safety and Fuel Efficiency. *Journal of Law and Economics,* **38:** 79-105.

Drummond, M.F., O'Brien B.J., Stoddart, G.L., Torrance, G.W. *Methods for the economic evaluation of health care programmes* (3rd edn.). New York: Oxford University Press, 2005.

Dyer, J.S., Sarin, R.K. 1979. Relative Risk Aversion. *Management Sciences,* **28:** 810-822.

European Commission. 2001. Recommended interim values for the value of preventing a fatality in DG Environment cost benefit analysis. http://europa.eu.int/comm/environment/enveco/others/recommended_interim_values.pdf

Featherstonhaugh, D., P. Slovic, S. M. Johnson, and J. Friedrich. 1997. "Insensitivity to the Value of Human Life: A Study of Psychological Numbing," *Journal of Risk and Uncertainty* **14,** 283–300.

Fernandez, J., Gonzaelz, M., Gascon, T., Moreno, J., Bidagor, J., Lagos, M., Rivas, F. 2010. Differences between willingness to pay and willingness to accept for visits by a family physician: A contingent valuation study. *BMC Public Health,* **10:** 236.

Fischer, G.W: Willingness to pay for probabilistic improvements in health status. A phsychological perspective. In Mushkin, S.J., Dunlop, D.W., eds: Health, What is it Worth? New York: Pregamon, 1979, pp.167-200.

Flowers CR, Garber AM, Bergen MR, Lenert LA. 1997. Willingness-to-pay utility assessment: feasibility of use in normative patient decision support systems, Journal of the American Medical Informatic Association Proceedings: AMIA Annual Fall Symposium: 223–227.

Foster, C.D., Beesley, M.E. 1963. Estimating the Social Benefit of Constructing an Underground Railway in London. *Journal of the Royal Statistical Society,* **126** (1).

Fox-Rushby, J. 1993. Appraising the use of contingent valuation: A Note in response. *Health Economics,* **2:** 361-362.

Frew, E.J. 2003. *Willingness to pay for colorectal cancer screening: a comparison of elicitation formats*. PhD Thesis. University of Nottingham.

Frew, E.J., Wolstenholme, J.L., Whynes, D.K. 2001. Willingness to pay for colorectal cancer screening. *European Journal of cancer,* **37:** 1746-51.

Frew, E.J., Wolstenholme, J.L., Whynes, D.K. 2004. Comparing willingness-to-pay: bidding game format versus open-ended and payment scale formats. *Health Policy,* **68:** 289-98.

Frew, E.J., Whynes, D.K., Wolstenholme, J.L. 2003. Eliciting Willingness to Pay: Comparing Closed-Ended with Open-Ended and Payment Scale Formats. *Medical Decision Making,* **23:** 150-159.

Froberg, D.G., Kane, R.L. 1989. Methodology for measuring health state preferences-II: Scaling Methods. *Journal of Clinical Epidemiology* **42** (5): 459-471

Fryback, D.G., Dasbach, E.J., Klein, R., Klein, B.E., Dorn, N., Peterson, K., Martin, P.A. 1993. The Beaver Dam Health Outcomes Study: initial catalogue of health-state quality factors. *Medical Decision Making,* **13** (2): 89-102

Gabriel, S.E., Campion, M.E., O'Fallon, W.M. 1994. A cost utility analysis of misoprostol prophylaxis for rheumatoid arthritis patients receiving non-steroidal anti-inflammatory drugs. *Arthritis and Rheumatology,* **37:** 333-41.

Gafni, A. 1991. Willingness-to-pay as a measure of benefits. Relevant questions in the context of public decision making about health care programs. *Medical Care,* **29**: 1246–52.

Gage, B.F., Cardinalli, A.B., Owens, D.K. 1996. The effect of stroke and stroke prophylaxis with aspirin or warfarin on quality of life. *Archives of Internal medicine,* **156:** 1829-1836.

Gandjour, A., Gafni, A., 2010. The additive utility assumption of the QALY model revisited. *Journal of Health Economics*. **29,** 325–328.

Glasziou, P.P., Bromwich, S., Simes, R.J. for the AUS-TASK Group. 1994. Quality of life six months after myocardial infarction treated with thrombolytic therapy. *Medical Journal of Australia,* **161:** 532-6.

Great Britain HM Treasury. 2004. *The Green Book.* Great Britain HM Treasury: London.

Green, C., Brazier, J., Deverill, M. 2000. Valuing Health Related Quality of Life: A Review of Health State Valuation Techniques. *Pharmacoeconomics,* **17** (2): 151-165.

Groome, P.A., Hutchinson, T.A., Tousignat, P., Hanley, J.A. 1999. The repeatability of Three methods for measuring prospective patients' values in the context of treatment choice for end stage renal disease. *Journal of Clinical Epidemiology,* **52** (9): 849-860.

Grosse, R. 1972. Cost-benefit analysis of health services. *Annals of the American Academy of Political and Social Science,* **399:** 89.

Grutters, J.P.C., Anteunis, L.J.C., Chenault, M.N., Joore, M.A. 2009. Willingness to pay for a hearing aid: comparing the payment scale and open-ended question. *Journal of Evaluation in Clinical Practice,* **15:** 91-96.

Guria, J., Leung, J., Jones-Lee, M., Loomes, G. 2005. The Willingness to Accept Value of a Statistical Life Relative to the Willingness to Pay Value: Evidence and Policy Implications. *Environmental and Resource Economics,* **32:** 113-127.

Gyldmark, M., Morrison, G.C. 1993. Re-appraising the use of contingent valuation: A reply. *Health economics,* **2** (4): 363-365

Gyrd-Hansen, D .2002 Comparing the results of applying different methods of eliciting time preferencesfor health. *European Journal of Health Economics* **3**: 10–16

Gyrd-Hansen, D. 2003. Willingness to pay for a QALY. *Health Economics,* **112:** 1049-1060

Gyrd-Hansen, D. 2005. Willingness to pay for a QALY: Theoretical and Methodological Issues. *Pharmacoecoomics,***23** ***(5):*** *423-432.*

Hackl, F., Pruckner, G.J. 2005. Warm glow, free-riding and vehicle neutrality in a health-related contingent valuation study. *Health Economics,* **14** (3): 293-306.

Hall, J., Gerard, K., Salkeld, G., & Richardson, J. (1992). A cost utility analysis of mammography screening in Australia. *Social Science and Medicine*, **34**(9), 993–1004.

Hamermesh, D.S. 1978. Economic Aspects of Job Satisfaction. In O. Ashenfelter andW. Oates (eds.), *Essays in Labor Market Analysis*. New York: John Wiley and Sons, pp. 53–72.

Hammerschmidt, T., Zeitler, H., Leidl, R. 2003. Unexpected yes and no answering behaviour in the discrete choice approach to elicit willingness to pay: A methodological comparison with payment cards. *International Journal of Healthcare Finance Economics* **3** (3): 147-66.

Hammitt, J. K., Graham, J.D. 1999. Willingness to Pay for Health Protection: Inadequate Sensitivity to Probability? *Journal of Risk and Uncertainty* 18, 33–62.

Hammitt, J.K., Haninger, K. 2010. Valuing fatal risks to children and adults: Effects of disease, latency, and risk aversion. *Journal of Risk and Uncertainty,* **40** (1): 57-83.

Hammitt, J.K., Robinson, L.A. 2011. The Income Elasticity of the Value per Statistical Life: Transferring Estimates between High and Low Income Populations. *Journal of Benefit-Cost Analysis,* ***2*** *(1): 1-27.*

Handler, R.M., Hynes, L.M., Nease, R.F. 1997. Effect of locus of control and consideration of future consequences on time tradeoff utilities for current health. *Quality of Life Research,* **6:** 54-60.

Hanemann, W.M. 1985. Some Issues in Continuous and Discrete-Response Contingent Valuation Studies. *Northeastern Journal of Agricultural and Resource Economics,* **14** (1): 5-13.

Hanemann, W.M. 1991. Willingness to pay and willingness to accept: How much can they differ? *American Economic Review,* **81**(3): 635-647

Hara Associates Inc. (2000). *Benefit/Cost Analysis of Proposed Tobacco Products Information Regulations*. Prepared for Health Canada and Consulting and Audit Canada. Ottawa, Ontario, June 5, 2000.

Hawthorne G, Richardson J, Osborne R. The assessment of quality of life (AQoL) instrument: a psychometric measure of health related quality of life. *QoL Res* 1999; **8**: 209–224.

Heinzen, R.R., Bridges, J.F.P. 2008. Comparison of four contingent valuation methods to estimate the economic value of a pneumococcal vaccine in Bangladesh. *International Journal of Technology Assessment in Health Care,* **24** (4): 481-487.

Hicks, J.R. 1939. The foundation of welfare economics. *Economic Journal,* **49:** 696-712

Hicks, J.R. 1941. The rehabilitation of consumer's surplus. *Review of Economic Studies,* **8:** 108-115.

Hintermann, B., A. Alberini, and A. Markandya. 2010. "Estimating the Value of Safety with Labour Market Data: Are the Results Trustworthy?" *Applied Economics,* **42** (7–9): 1085–100.

Hirth, R.A., Chernew, M.E., Miller, E., Fendrick, M., Weisser, W.G. 2000. Willingness to pay for a quality adjusted life year: in search of a standard. *Medical Decision Making,* **20:** 332-342.

Hornberger JC, Redelmeier DA, Petersen J. 1992. Variability among methods to assess patients well-being and consequent effect on a cost-effectiveness analysis. *Journal of Clinical Epidemiology*; **45**: 505–512.

Hosmer, D.W., Lemeshow, S. 2000. *Applied Logistic Regression,* New York: Wiley.

House of Commons Health Select Committee. (2007). *First Report of Sessions 2007-8 on the National Institute for Health and Clinical Excellence.*
http://www.publications.parliament.uk/pa/cm200708/cmselect/cmhealth/27/2702.htm accessed 26 May 2008

House of Commons Health Committee: *National Institute for Clinical Excellence*. Second Report of Session 2001-02. The Stationary Office. 2002.

Hwang, H.-S., W.R. Reed, and C. Hubbard. 1992. "Compensating Wage Differentials and Unobserved Productivity." *The Journal of Political Economy* 100(4): 835–58.

Institute for Social and Economic Research (ISER). 2007. www.iser.essex.ac.uk/bhps.

Jacobs, R.J., Moleski, R.J., Meyerhoff, A.S. 2002. Valuation of Symptomatic Hepatitis A in Adults. Estimates based on Time Trade-Off and Willingness to Pay Measurement. *Pharmacoeconomics,* **20** (11): 739-747

Johannesson, M., Johansson, P-O. 1997. Quality of life and the WTP for an increase in life expectancy at an advanced age. *Journal of Public Economics,* **65**: 219-228

Johannesson M, Johansson P-O, Kristrom B, Borgquist L, Jonsson B. 1993. Willingness to pay for lipid lowering: a health production function approach. *Applied Economics*; **25**: 1023–31.

Johannesson, M., Johansson, P.-O., Kristrom, B. and Gerdtham, U.-G. 1993. Willingness to pay for antihypertensive therapy: further results. *Journal of Health Economics*; **12**: 95–108.

Johannesson M, Johnsson B, Borgquist L. 1991. Willingness to pay for antihypertensive therapy-results of a Swedish pilot study. *Journal of Health Economics*; 10: 461–74.

Johannesson M, Jonsson B, Karlson G. 1996. Outcome measurement in economic evaluation. *Journal of Health Economics*; **5**: 279–96.

Johannesson, M., Jonsson, B. 1991. Economic Evaluation in health care: is there a role for cost-benefit analysis? *Health Policy,* **17(1):** 1-23

Johannesson, M., Meltzer, D. 1998. Some reflections on cost-effectiveness analysis. *Health Economics,* **7** (1): 1-7

Johannesson M, O'Conor RM, Kobelt-Nguyen G, Mattiasson A. 1997. Willingness to pay for reduced incontinence symptoms. *British Journal of Urology*; 80: 557–62.

Johannesson M, Pliskin JS, Weinstein MC. 1994. A note on QALYs, time trade off, and discounting. *Medical Decision Making;* **14**: 188–193

Johansson, P-O. *An Introduction to Modern Welfare Economics.* Cambridge University Press, Cambridge, 1991.

Johnson, E.S., Sullivan, S.D., Mozaffari, E. 1996. A utility assessment of oral and intravenous gancilcovir for the maintenance treatment of AIDS related cytomegalovirus retinitis. *Pharmacoeconomics,* **10:** 623-9.

Johnson, F.R. 2005. Einstein on willingness to pay per QALY: is there a better way? *Medical Decision Making,* **25** (6): 607-8.

Johnson, F.R., Desvousges, W.H., Ruby, M., Stieb, D. and De Civita, P. 1998. Eliciting stated health preferences: an application to willingness to pay for longevity. *Medical Decision Making,* **18:** S57-S67.

Jones-Lee, M. W. 1974. The Value of Changes in the Probability of Death or Injury, *Journal of Political Economy* **82**, 835–849.

Jones-Lee, M.W., M. Hammerton, and P.R. Phillips. 1985. The Value of Safety: Results of a National Sample Survey. *Economic Journal,* **95**: 49–72.

Jones-Lee, M.W., Loomes, G., Jones, S., Rowlatt, P., Spackman, M., Jones, S. 1998. *Valuation of Deaths from Air Pollution.* NERA and CASPAR. Report prepared for the Department of Environment, Transport and the Regions and the Department of Trade and Industry, London.

Kahneman, D.J., Knetsch, J., 1992. Valuing public goods: the purchase of moral satisfaction. *Journal of Environmental Economics and Management* **22**, 57–70.

Kahneman, D., Tversky, A. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica,* **47**(2): 263-291

Kaldor, N. 1939. Welfare propositions of economics and interpersonal comparisons of utility. *Economic Journal,* **49:** 549-52

Kanninen BK. Bias in discrete response contingent valuation. 1995. *Journal of Environmental Economics and Management*; 28: 114–25.

Kartman B, Andersson F, Johannesson M. 1996. Willingness to pay for reductions in angina pectoris attacks. *Medical Decision Making*; 16: 248–53.

Kartman B, Stalhammar N-O, Johannesson M. 1996. Valuation of health changes with the contingent valuation method: a test of scope and question order effects. Health Economics; 5: 531–41.

Keeney R, Raiffa H. Decisions with Multiple Objectives. Wiley: New York, 1976.

King Jr., J.T., Tsevat, J., Lave, J.R., Roberts, M.S. 2005. Willingness to pay for a quality adjusted life year implications for societal health care resource allocation. *Medical Decision Making,* **25** (6): 667-677.

Kirsch, J., & McGuire, A. (2000). Establishing health state valuations for disease specific states: an example from heart disease. *Journal of Health Economics,* **9**(2), 149–158.

Klarman, H. 1967. Present status of cost-benefit analysis in the health field. *American Journal of Public Health,* **57:** 1948.

Klarman, H. 1974. Application of cost-benefit analysis to heath systems technology. *Journal of Occupational Medicine,* **16**: 172.

Klose, T. 1999. The contingent valuation method in health care. *Health Policy,* **47:** 97-123.

Kniesner, T.J., Leeth, J.D., 1991. Compensating wage differentials for fatal injury risk in Australia, Japan and the United States. *Journal of Risk and Uncertainty*, 4 (1), 75–90.

Kniesner, T.J., W.K. Viscusi, C. Woock, and J.P. Ziliak. 2012. The Value of Statistical Life: Evidence from Panel Data. *The Review of Economics and Statistics,* **94** (1): 74-87

Kniesner, T.J., W.K. Viscusi, and J.P. Ziliak. 2010. Policy Relevant Heterogeneity in the Value of Statistical Life: New Evidence from Panel Data Quantile Regressions. *Journal of Risk and Uncertainty* 40(1): 15–31.

Koch, G.G. 1982. Intraclass correlation coefficient. In: Kotz, S., Johnson, N.L. (eds.), *Encyclopaedia of Statistical Sciences*, **4:** 212-217.

Kochi I, Hubbell B, Kramer R. 2006. An Empirical Bayes Approach to Combining and Comparing Estimates of the Value of a Statistical Life for Environmental Policy Analysis. *Environmental and Resource Economics*, **34** (3): 385-406

Kontodimopoulos, N., Niakas, D. 2006. Overcoming inherent problems of preference-based techniques for measuring health benefits: An empirical study in the context of kidney transplantation. *BMC Health Services Research,* **6:** 3.

Koopmanschap, M., Rutten, F.H.H. 1993. Indirect costs in economic studies. *Pharmacoeconomics*, **4**: 446-54

Koopmanschap, M.A., van Ineveld, B. 1992. Towards a new approach for estimating indirect costs of disease. *Social Science and Medicine,* **34**: 1005-10

Koopmanschap, M., Rutten, F., van Ineveld, B., Van Roijen, L. 1995. The friction cost method of measuring the indirect costs of disease. *Journal of Health Economics*, **14**: 171-89

Krabbe, P.F.M., Bonsel, G.J. 1998. Sequence effects, health profiles, and the QALY model: in search of realistic modelling. *Medical Decision Making*; **18**(2):178–86.

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

Kreibich, D.N., Vaz, M., Bourne, R.B., Rorabeck, C.H., Kim, P., Hardie, R., Kramer, J., Kirkley, A. 1996. What is the best way of assessing outcome after total knee replacement? *Clinical Outcomes and Related Research,* **331:** 221-225.

Krumins, P.E., Fihn, S.D., Kent, D.L. 1988. Symptom severity and patients' values in the decision to perform a transneutral resection of the prostate. *Medical Decision Making,* **8:** 1-8.

Krupnick, A. 2007. Mortality-risk Valuation and Age: Stated Preference Evidence. *Review of Environmental Economics and Policy,* **1** (2): 261–282

Krupnick, A., M. Cropper, A. Alberini, N. Simon, B. O'Brien, and R. Goeree. 2002. "Age, Health and the Willingness to Pay for Mortality Risk Reductions: A Contingent Valuation Survey of Ontario Residents." *Journal of Risk and Uncertainty* **24**(2): 161–75.

Lakhani N: The Price of life: it was £20,000. Now an NHS agency recalculates. Independent on Sunday 2008, 18-19.

Lanoie, P., Pedro, C., Latour, R., 1995. The value of a statistical life: a comparison of two approaches. *Journal of Risk and Uncertainty* **10**, 235–257.

Lattimore PM, Baker JR, Witte AD. 1992. The Influence of probability on risky choice. Journal of Economic Behaviour and Organisations, **17**: 377–400.

Layard, R., Mayraz, G., Nickell, S. 2008. The marginal utility of income. *Journal of Public Economics,* **92** (8): 1846 – 1857.

Lee SJ, Neumann PJ, Churchill WH, Cannon ME, Weinstein MC, Johannesson M. 1997. Patients' willingness to pay for autologous blood donation. *Journal of Health Policy*; 40: 1–12.

Leigh, J.P. 1981. Compensating Wages for Occupational Injuries and Diseases, *Social Science Quarterly* **62**(4), 772–778.

Leigh, J. 1995. Compensating Wages, Value of a Statistical Life, and Inter-Industry Differentials. *Journal of Environmental Economics and Management* **28**(1): 83–97.

Lenert LA, Cher DJ, Goldstein MK, Bergen MR, Garber AM. 1998. The effect of search procedures on utility elicitations. *Journal of Medical Decision Making*. **18**: 76–83.

Lenert, LA, Sherbourne CD, Reyna V. 2001. Utility elicitation using single-item questions compared with a computerized interview. *Journal of Medical Decision Making*. **21**: 97-104.

Levenson H. Activism and powerful others. 1974. Distinctions within the concept of internal-external control. *Psychological Reports*; **38**: 377-383.

Lieu, T.A, Thomas Ray, G., Ortega-Sanchez, I.R., Kleinman, K., Rusinak, D., Prosser, L.A. 2009. Willingness to pay for a QALY based on community member and patient preferences for temporary health states associated with herpes zoster. *Pharmacoeconomics,* **27** (12): 1005-1016.

Lindhjem, H.S., Navrud, S., Braathen, N.A. 2010. Valuing Lives Saved from Environmental, Transport, and Health Policies: A Meta-Analysis of Stated Preference Studies. Prepared for the Organization for Economic Co-operation and Development. ENV/EPOC/WPNEP(2008)10/FINAL.

Lindholm, L., Rosen, M., Hellsten, G. 1994. Are people willing to pay for a community-based preventive program. *International Journal of Technology Assessment in Health Care,* **10:** 317-324.

Loehman ET, Berg SV, Arroyo AA, Hedinger RA, Schwartz JM, Shaw ME, Fahien RW, De VH, Fishe RP, Rio DE, Rossley WF, Green AES. 1979. Distribution analysis of regional benefits and cost of air quality control. *Journal of Environmental Economics and Management*. **6**: 222–43.

Long, J.S., Freese, J. 2006. *Regression Models for Categorical Dependent Variables Using Stata.* 2nd ed. College Station, TX: Stata Press.

Luce, R.D., Tukey, J.W. 1964. Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology,* **1** (1): 1-27.

MacKeigan LD, O'Brien BJ, Oh PI. 1999. Holistic versus composite preferences for lifetime treatment sequences for type 2 diabetes. Medical Decision Making. **19**(2):113–21.

Mann, C.C. Can meta-analysis make policy? *Science,* **266** (5187): 960-962.

Martin, A. J., Glasziou, P. P., Simes, R. J., & 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.

Marin, A., Psacharopoulos, G. 1982. The reward for risk in the labour market: Evidence from the United Kingdom and reconciliation with other studies. *Journal of Political Economy,* **90:** 827-53.

Mason, H., Baker, R., Donaldson, C. 2008. Willingness to pay for a QALY: past, present and future. *Expert Review of Pharmacoeconomics and Outcomes Research,* **8**(6): 575-582.

Mason, H., Donaldson, C. Does QALY type matter in estimating willingness to pay for health gains? Paper presented to the Health Economists' Study Group, University of Birmingham, January 2007.

McCabe, C., Claxton, K., Culyer, A.J. 2008. The NICE Cost-Effectiveness Threshold: What is is and what that means. *Pharmacoeconomics,* **26** (9): 733-744.

McNamee, P., Ternent, L., Gbangou, A., Newlands, D. 2010. A game of two halves? Incentive incompatibility, starting point bias and the bidding game contingent valuation method. *Health Economics,* **19:** 75-87.

McNeil BJ, Weichselbaum R, Pauker SG. 1978. Fallacy of the five-year survival in lung cancer. *New England Journal of Medicine*, **299**: 1397–1401.

Mcneil BJ, Weichselbaum R, Pauker SG. 1981. Speech and survival: trade-offs between quality and quantity of life in laryngeal cancer. *New England Journal of Medicine*, 305: 982–7.

Mehrez A, Gafni A. 1990. Evaluating health related quality of life: an indifference curve interpretation for the time trade-off technique. *Social Science Medicine*, **31**(11): 1281–1283.

Miedzybrodzka Z, Shackley P, Donaldson C, Abdalla M. 1994. Counting the benefits of screening: a pilot study of willingness to pay for cystic fibrosis carrier screening. *Journal of Medical Screening*; **1**: 82–3.

Miedzybrodzka Z, Semper J, Shackley P, Abdalla M, Donaldson C. 1995. Stepwise or couple antenatal screening for cystic fibrosis?: women's preferences and willingness to pay. *Journal of Medical Genetics*, **32**: 282–3.

Miller, T.R. 1990. The Plausible Range for the Value of Life – Red Herrings among the Mackerel. *Journal of Forensic Economics*, **3** (3): 17-39.

Mishan, E.J. Cost-benefit analysis. New York: Praeger, 1971.

Mishan, E.J. 1977. The plain truth about consumer surplus. *Zeitschrift fur Nationalokonomie*, **37**: 1-24

Mishan, E.J. Cost-benefit analysis: an informal introduction, 4[t] Edition. Unwin, Hyman, London, 1988.

Mitchell, R.C., Carson, R.T. 1984. *A contingent valuation estimate of national freshwater benefits: technical report to the U.S. environmental protection agency*. Resources for the Future, Washington DC.

Mitchell, R.C., Carson, R.T. *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Resources for the Future, Washington DC, 1989.

Miyamoto JM, Eraker SA. 1988. A multiplicativemodel of the utility of survival duration and health quality. Journal of Experimental Psychology, **117**(1):3–20.

Molzahn, A.E., Northcott, H.C., hayduk, L. 1996. Quality of life of patients with end stage renal disease: a structured equation model. *Quality of Life Research*, **5**: 426-32.

Mooney, G.H. 1980. Cost benefit analysis and medical ethics. *Journal of Medical Ethics*, **6**: 177-179

Mooney, G. 1992. Economics, medicine and health care. Wheatsheaf, Hemel Hempstead.

Morrison, G.C. and Gyldmark, M. 1992. Appraising the use of contingent valuation. *Journal of Health Economics*; **1**: 233–43.

Mrozek, J.R., Taylor, L.O. 2002. What determines the Value of Life? A meta-analysis. *Journal of Policy Analysis and Management,* **21** (2): 253-270.

Muller A, Reutzel TJ. 1984. Willingness to pay for reduction in fatality risk: an exploratory survey. *American Journal of Public Health*, **74**: 804–12.

National Institute for Clinical Excellence. *NICE guide to the methods of technology appraisal.* NICE, UK, 2008.

National Institute for Clinical Excellence. *Appraising life-extending, end of life treatments*. NICE, UK, 2009.

Neumann PJ, Johannesson M. 1994. The willingness to pay for in vitro fertilization: a pilot study using contingent valuation. *Medical Care*, **32**: 686–99.

Nord, E. 1992. Methods for Quality Adjustment of Life Years. *Social science and Medicine,* **34** (5): 559-569.

Norinder, A., Hjalte, K., Persson, U. 2001. Scope and scale insensitivities in a contingent valuation study of risk reductions. *Health Policy,* **57** (2): 141-153

O'Brien, B. and Gafni, A. When do the dollars make sense? Toward a conceptual framework for contingent valuation studies in health care. *Medical Decision Making* 1996; **16**: 288–99.

O'Brien, B.J., Gertsen, K., Willan, A.R., Faulkner, L.A. 2002. Is there a kink in consumers' threshold value for cost-effectiveness in health care. *Health Economics,* **11:** 175-180.

O'Brien, B.J., Goeree, R., Gafni, A., Torrance, G.W., Pauly, M.V., Erder, H., Rusthoven, J., Weeks, J., Cahill, M., Lamont, B. 1998. Assessing the value of a new pharmaceutical: a feasibility study of contingent valuation in managed care. *Medical Care,* **36:** 370-84.

O'Brien, B.J., Viramontes, J.L. 1994. Willingness to pay: a valid and reliable measure of health state preference? *Medical Decision Making,* **14:** 289-297

O'Connor, A.M., Boyd, N.F., Warde, P., Stolbach, L., Till, J.E. 1986. Eliciting preferences for alternative drug therapies in oncology: influence of treatment outcome description, elicitiation techninque and treatment experience on preferences. *Journal of Chronic Diseases,* **40** (8): 811-818.

OECD (2010) *Valuation of Environmental Health Risks for Children: Draft Report,* Environment Directorate, Environment Policy Committee, Working Party on National Environmental Policies, ENV/EPOC/WPNEP(2008)9/FINAL

Olsen, J.A. 1994. Persons vs years: two ways of eliciting implicit weights. *Journal of Health Economics*, **3**: 39–46

Olsen JA, Donaldson C. Willingness to pay for public sector health care programmes in Northern Norway. University of Aberdeen: HERU Discussion Paper 05:93.

Olsen, J.A., Donaldson, C. 1998. Helicopters, hearts and hips: using willingness to pay to set priorities for public sector health care programmes. *Social Science and Medicine*, **46**: (1): 1-12

Olsen, J.A., Donaldson, C., Pereira, J. 2004. The insensitivity of 'willingness-to-pay' to the size of the good: New evidence for health care. *Journal of Economic Psychology*, **25**: 445-460

Olsen, J.A., Smith, R. 2001. Theory versus practice: A review of 'willingness to pay' in health and health care. *Health Economics*, **10**(1): 39-52

Onwujekwe, O., Fox-Rushby, J., Hanson, K. 2008. Construct Validity of the Bidding Game, Binary with Follow-up, and a novel structured haggling question format in determining willingness to pay for insecticide-treated mosquito nets. *Medical Decision Making*, **28**: 90-101.

Onwujekwe, O., Hanson, K., Fox-Rushby, J. 2005. Do divergences between stated and actual willingness to pay signify the existence of bias in contingent valuation surveys? *Social Science and Medicine*, **60**: 525-536

Patrick *et al.* D.L., Bush, J.W., Chen, M.M. 1973. Methods for measuring levels of well-being for a health status index. *Health Services research*, **8**: 228-245.

Patrick, D.L., Starks, H.E., Cain, K.C. 1994. Measuring preferences for health states worse than death. *Medical decision Making*, **14**: 9-18.

Pauly, M.V., Valuing health care benefits in money terms. In: *Valuing Health Care*, Sloan, F.A. (ed.). Cambridge University press: Cambridge, 1995.

Phelps, C.E., Mushlin, A. On the (near) equivalence of cost-effectiveness and cost-benefit analyses. University of Rochester School of Medicine and Dentistry Press, 1991.

Phillips KA, Homan RK, Luft HS, Hiatt PH, Olson KR, Kearney TE, Heard SE. 1997. Willingness to pay for poison control centers. *Journal of Health Economics*, **19**: 343–57.

Pliskin, J.S., Shepard, D.S., Weinstein, M.C. 1980. Utility functions for life years and health status. *Operations Research*, **28**: 206-24

Prades, J.L.P, Loomes, G., Brey, R. 2009. Trying to estimate a Monetary Value for the QALY. *Journal of Health Economics*, **28**: 553-562

Pratt, J. W. and R. J. Zeckhauser. 1996. Willingness to Pay and the Distribution of Risk and Wealth, *Journal of Political Economy* **104**, 747–763.

Pregibon, D. 1981. Logistic Regression Diagnostics. *The Annals of Statistics,* **9** (4): 705-724.

Ratcliffe, J. 2000. The use of conjoint analysis to elicit willingness to pay. Proceed with caution? *International Journal of Technology Assessment in Health Care,* **16:** 270-90.

Ratcliffe, J., Buxton, M., McGarry, T., Sheldon, R., Chancellor, J. 2004. Patients' preferences for characteristics associated with treatments for arthritis. *Rheumatology,* **43:** 337-45.

Rawlins, M.D., Culyer, A.J. 2004. National Institute for Clinical Excellence and its value judgements. *BMJ,* **329**:224-227

Read JL, Quinn DM, Berrick HV, Fineberg HV, Weinstein ML. 1984. Preferences for health outcomes: comparison of assessment methods. *Medical Decision Making*; **4**: 315–329.

Reed, W.W., Herbers, J.E., Noel, L. 1993. Cholesterol lowering therapy: what patients expect in return. *Journal of General Internal Medicine,* **8:** 591-6.

Rice, D., Cooper, B. 1967. The economic value of human life. *American Journal of Public Health,* **57**: 1954-66

Richards, M. Improving access to medicines for NHS patients. Department of Health, London; 2008.

Richardson, J. 1994. Cost utility analysis: What should be measured? *Social Science and Medicine,* **39** (1): 7-21.

Richardson J, Hall J, Salkeld G. 1996. The measurement of utility in multiphase health states. *International Journal of Technological Assessments in Health Care,* **12** (1): 151–62

Robinson, A., Dolan, P., Williams, A. 1997. Valuing Health Status using VAS and TTO: What lies behind the numbers? *Social Science and Medicine,* **45** (8): 1289-1297.

Robinson, A., Gyrd-Hansen, D., Bacon, P., Baker, R., Pennington, M., Donaldson, C., and the EuroVaQ Team. Estimating a WTP-based value of a QALY: the 'chained approach'. Paper presented to the 3rd Joint CES/HESG Meeting, Aix en Provence, 11-13th January 2012.

Rosen, S. 1988. The value of changes in life expectancy. *Journal of Risk and Uncertainty,* **1** (3): 285-304

Ross CK, Steward CA, Sinacore JM. 1995. A comparative study of seven measures of patient satisfaction. *Medical Care, 33*: 392–406.

Ryan M. Using the closed ended willingness to pay technique to establish arguments in the infertile person's utility function. University of Aberdeen, UK: HERU Discussion Paper 07:95.

Ryan M. 1996. Using willingness to pay to assess the benefits of assisted reproductive techniques. *Journal of Health Economics*, **5**: 543–58

Ryan, M., Gerard, K. 2003. Using discrete choice experiment to value health care programmes: current practice and future research reflections. *Applied Health Economics and Health Policy,* **2:** 55-64.

Ryan, M., Scott, D.A., Donaldson, C. 2004. Valuing health care using willingness to pay: a comparison of the payment card and dichotomous choice methods. *Journal of Health Economics,* **23:** 237-58.

Ryan, M., Scott, D.A., Reeves, C., Bate, A., van Teijlingen, E.R., Russell, E.M., Napper, M., Robb, C.M. 2001. Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technology Assessment,* **5** (5): 1-186.

Ryan, M., Watson, V., Amaya-Amaya, M. 2003. Methodological issues in the monetary valuation of benefits in healthcare. *Expert Review of Pharmacoeconomics and Outcomes Research,* **3** (6): 717-727.

Sach T, Smith RD, Whynes D. 2007. A 'league table' of contingent valuation results for pharmaceutical interventions: a hard pill to swallow? *PharmacoEconomics* **25**: 107–127

Sackett,D., Torrance,G. 1978. The utility of different health states as perceived by the general public. *Journal of Chronic Disorders*. **31**: 697–704.

Sandy, R., Elliott, R.F. 1996. Unions and risk: their impact on the compensation for fatal risk. *Economica* **63** (250), 291–309.

Scott, A., Watson, M.S., Ross, S. 2003. Eliciting preferences of the community for out-of-hours care provided by general practitioners: a stated preference discrete choice experiment. *Social Science and Medicine,* **50:** 804-814.

Sculpher, M.J., Buxton, M.J. 1993. The episode-free day as a composite measure of effectiveness. *Pharmacoeconomics,* **4:** 345-52.

Sculpher, M.J., Bryan, S., Fry, P., de Winter, P., Payne, H., Emberton, M. 2004. Patients' preferences for the management of non-metastatic prostate cancer: discrete choice experiment. *British Medical Journal,* **328:** 382.

Shiell, A., Hawe, P. 2006. Test-retest reliability of willingness to pay. *European Journal of Health Economics,* **7:** 176-181.

Shiroiwa, T., Sung, Y.K., Fukuda, T., Lang, H., Bae, S., Tsutani, K. 2010. International Survey on willingness to pay (WTP) for one additional QALY gained: what is the threshold of cost effectiveness. *Health Economics,* **19:** 422-437.

Siebert,W.S.,Wei, X., 1994. Compensating wage differentials for workplace accidents: evidence for union and non-union workers in the UK. *Journal of Risk and Uncertainty,* **9**, 61–76.

Sinclair, J., Torrance, G., Boyle, M., Horwood, S., Saigal, S. and Sackett, D. 1981. Evaluation of neonatal intensive care programs. *New England Journal of Medicine,* **305:** 489-94

Skjoldborg, U.S., Gyrd-Hansen, D. 2003. Conjoint Analysis. The cost variable: an Achille's heel? *Health Economics,* **12:** 479-91

Sox HC, Blatt MA, Hig gins MC, Mar ton KI. 1986. *Medical decision making.* Butterworths: Boston Press.

Smith, R.D. 2000. The discrete choice willingness to pay question format in Health Economics: should we adopt environmental guidelines? *Medical Decision Making,* **20** (2): 194-206.

Smith, R.D. 2001. The relative sensitivity of willingness to pay and time trade off to change in health status: an empirical investigation. *Health Economics,* **10:** 487-497.

Smith, R.D. 2003. Construction of the contingent valuation market in health care: a critical assessment. *Health Economics,* **12:** 609-628

Smith, R.D. 2004. The reliability of Willingness to Pay for Changes in Health Status. *Applied Health Economics and Health Policy,* **3** (1): 35-38

Smith, R.D. 2005. Sensitivity to scale in contingent valuation the importance of the budget constraint. *Journal of Health Economics,* **24** 515-529

Smith, R.D. 2006. It's not what you do it's the way that you do it: The effect of different payment card formats and survey administration on willingness to pay for health gain. *Health Economics,* **15** (3): 281-293

Smith, R.D. 2007. The Relationship between reliability and size of willingness to pay values: A qualitative insight. *Health Economics,* **16** (2) 211-216

Smith, R.D., Richardson, J. 2005. Can we estimate the 'social' value of a QALY? Four core issues to resolve. *Health Policy,* **74:** 77-84

Smith, R.D., Sach, T.H. 2009. Contingent Valuation: (Still) on the road to nowhere? *Health Economics,* **18:** 836-866

Smith, R.D., Sach, T.H. 2010. Contingent Valuation: what needs to be done? *Health Economics, Policy and Law,* **5:** 91-111.

Spencer A. 2001. The TTO method and procedural invariance. *Health Economics,* 12: 655–668.

Spencer, A. 2004. The implications of linking questions within the SG and TTO methods. *Health Economics,* **13:** 807-818.

Stalhammer, N.O. 1996. An empirical note on willingness-to-pay and starting point bias. *Medical Decision Making,* **16:** 242-7.

Stalmeier, P. F. M., Chapman, G. B., de Boer, A. G. E. M., & van Lanschot, J. J. B. 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**(4), 488–496.

Stalmeier, P.F.M., Bezembinder, T.G.G., Unic, I.J. 1996. Proportional heuristics in time tradeoff and conjoint measurement. *Medical Decision Making,* **16** (1): 36–44.

Stalmeier, P.F.M., Lamers, L.M., Busschbach, J.V., Krabbe, P. 2007. On the assessment of preferences for health and duration: maximal endurable time and better than dead preferences. *Medical Care,* **45** (9): 835-841.

Stalmeier P,Wakker P, Bezembinder T. 1997. Preference reversals: violations of unidimensional procedure invariance. *Journal of Experimental Psychology Hum. Percept Perform,***23** (94): 1196–205.

Stavem K, Kris tiansen IS, Olsen JA. 2002. Association of time preference for health with age and disease severity. *European Journal of Health Economics* **3**: 120–124

Stewart, J.M., O'Shea, E., Donaldson, C., Shackley, P. 2002. Do ordering effects matter in willingness-to-pay studies of health care? *Journal of Health Economics,* **21:** 585-599.

Stiggelbout, A. M., Kiebert, G. M., Kievit, J., Leer, J. W., Habbema, J. D., & De Haes, J. C. 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 AM, Kiebert GM, Kievit J, Leer JWH, Stoter G, de Haes JCJM. 1994. Utility assessment in cancer patients: adjustment of time trade-off scores for the utility of life years and comparison with standard gamble scores. *Medical Decision Making,* **14**: 82–90.

Strand, J. 2004. *Public- and Private-good Values of Statistical Lives: Results from a Combined Choice-experiment and Contingent-valuation Survey.* Unpublished manuscript, Department of Economics, University of Oslo.

Sugden, R. 1999. 'Public Goods and Contingent Valuation', Chapter 5, pp. 131–152, and

'Alternatives to the Neo-classical Theory of Choice', Chapter 6, pp. 152–181 in I. Bateman I and K. Willis, eds., Valuing Environmental Preferences. Oxford: Oxford University Press.

Sutherland HJ, Llewellyn-Thomas H, Boyd NF, Till JE. 1982. Attitude toward quality of survival: the concept of maximal endurable time. *Medical Decision Making*, **2**: 299–309.

Telser, H., Zweifel, P. 2007. Validity of discrete-choice experiments for health risk reduction. *Applied Economics,* **39** (1): 69-78

Thaler, R. 1985. Mental accounting and consumer choice. *Marketing Science,* **4** (3): 199-214.

Thaler, R., Rosen, S., 1975. The Value of Saving a Life: Evidence from the Labor Market. In N.E. Terleckyj (ed.), *Household Production and Consumption.* New York: Columbia University Press, pp. 265-300.

Thompson MS, Read JL, Liang M. 1984. Feasibility of willingness-to-pay measurement in chronic arthritis. *Medical Decision Making*, **4**: 195–215

Thompson MS, Read JL, Liang M. *Willingness-to-pay concepts for societal decisions in health. In: Kane RL, Kane RA, editors*. Values and long-term care. Toronto: Lexington, 1982.

Tilling, C.J., Devlin, N., Tsuchiya, A., Buckingham, K. 2010. Protocols for Time Trade-off Valuations of Health States Worse than Dead: A Literature Review. *Medical Decision Making,* **30** (5): 610-619

Tilling, C.J., Krol, M., Tsuchiya, A., Brazier, J.E., van Exel, J., Brouwer, W.B.F. 2012. Does the EQ-5D reflect lost earnings? *Pharmacoeconomics,* **30** (1): 47-61

Tolley, G., Kenkel, D. and Fabian, R. 1994, State of the Art Health Values. Tolley, G., Kenkel, D. and Fabian, R.(eds), In: *Valuing Health for Public Policy, an Economic Approach*. University of Chicago Press, Chicago. 323–44.

Tolley G, Kenkel D, Fabian R, editors. Valuing health for policy: an economic approach. Chicago, IL: University of Chicago Press, 1994.

Torrance, G.W., 1976. Social preferences for health states: an empirical evaluation of three measurement techniques. *Socioeconomic Planning Sciences,* **10:** 129-136.

Torrance, G.W. 1986. Measurement of health-state utilities for economic appraisal: a review. *Journal of Health Economics,* **5:** 1-30.

Torrance, G.W. 1987. Utility approach to measuring health-related quality of life. *Journal of Chronic Disease,* **40:** 593-603.

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

Torrance GW, Feeny D. 1989. Utilities and quality adjusted life years. *International Journal of Technological Assessments in Health Care*; **5**: 559–575.

Torrance, G.W., Thomas, W.H., Sackett, D.L. 1972. A utility maximization model for evaluation of health care programs. *Health Services Research,* **7:** 118-133.

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. *Medical Decision Making,* **25:** 460-467.

Tversky A, Kahneman D. 1992. Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*; **5**: 297–323.

Tversky A, Sattath S, Slovic P. 1988. Contingent weighting in judgment and choice. *Psychology Review,* **95**: 371–384.

Unic I, Stalmeier P, Vefhoef LvDW. 1998. Assessment of the time-trade-off values for prophylactic mastectomy of women with a suspected genetic predisposition to breast cancer. *Medical Decision Making,* **18** (3): 268–77

U.S. Environmental Protection Agency (EPA). 1999. *The Benefits and Costs of the Clean Air Act, 1990 to 2010.* EPA 410-R-99-001.

U.S. Environmental Protection Agency (EPA). 2006. *Regulatory Impact Analysis for the National Ambient Air Quality Standards for Particle Pollution.*

U.S. Office of Management and Budget. (1996). *Economic Analysis of Federal Regulations Under Executive Order 12866.* Washington, DC: The White House. Internet: http://www.whitehouse.gov/omb/inforeg/riaguide.html.

Van der Donk, J., Levendag, P.C., Kuijpers A.J., Roest, F.H., Habbema, J.D., Meeuwis, C.A., Schmitz, P.I. 1995. Patient participation in clinical decision making for treatment of T3 laryngeal cancer: a comparison of state and process utilities. *Journal of Clinical Oncology,* **13:** 2369-78.

Van der Pol, M., & Roux, L. (2005). Time preference bias in time trade-off. European Journal of Health Economics, **6**, 107–111.

Van Nooten, F.E., Koolman, X., Brouwer, W.B.F. 2009. The influence of subjective life expectancy on health state valuations using a 10 year TTO. *Health Economics,* **18**(5): 549-558

Viscusi, W.K. 1978. Wealth effects and earnings premiums for Job Hazards. *Review of Economics and Statistics,* **60** (3): 408-416.

Viscusi, W.K. 1979. *Employment Hazards: An Investigation of Market Performance.* Cambridge, MA: Harvard University Press.

Viscusi, W.K. 1980. Union, Labor Market Structure, and the Welfare Implications of the Quality of Work. *Journal of Labor Research* **1**(1), 175–192.

Viscusi, W.K. 1993. The Value of Risks to Life and Health. *Journal of Economic Literature,* **8** (1): 5-17.

Viscusi, W.K. 2004. The Value of Life: Estimates with Risks by Occupation and Industry. *EconomicInquiry* 42(1): 29–48.

Viscusi WK. 2009. The Devaluation of Life. *Regulation and Governance.* **3** (2): 103-27

Viscusi, W.K., Aldy, J.E. 2003. The Value of a Statistical Life: A critical review of market estimates throughout the world. *The Journal of Risk and Uncertainty,* **27** (1): 5-76.

Verhoef LCG, de Haan AFJ, van Daal WAJ. 1994. Risk attitude in gambles with years of life: empirical support for prospect theory. *Medical Decision Making,* **14**: 194–200.

von Neumann, J., Morgenstern, O. 1944. *Theory of games and economic behaviour.* Princeton University Press, Princeton, New Jersey.

Wakker PP, Deneffe D. 1996. Eliciting von Neumann– Morgenstern utilities when probabilities are distorted or unknown. *Management Science* **42**: 1131–1150.

Walraven G. 1996. Willingness to pay for district hospital services in rural Tanzania. Health Policy and Planning, **11**: 428–37.

Watson, V., Ryan, M. 2004. Testing Contingent Valuation Responses for Payment Vehicle Effects: Taxation v Charitable Donations. Paper presented at Health Economics Study Group Meeting, Galsgow.

Wee, H., Li, S., Xie, F., Zhang, X., Luo, N., Feeney, D., Cheung, Y., Machin, D., Fong, K., Thumboo, J. 2008. Validity, feasibility and acceptability of time trade-off and standard gamable assessments in health valuation studies: a study in a multi-ethnic Asian population in Singapore. *Value in Health,* **11** (1): S3-S10.

Weinstein, M.C., Fineberg, H.V. 1980. *Clinical Decision Analysis.* Saunders, Philadelphia.

Weinstein, M. C., D. S. Shepard, and J. S. Pliskin. 1980. The Economic Value of Changing Mortality Probabilities: A Decision-Theoretic Approach, *Quarterly Journal of Economics* **94**, 373–396.

Weinstein, M.C., Stason, W.B. 1977. Foundations of cost-effectiveness analysis for health and medical practices. *New England Journal of Medicine,* **296:** 716-721.

Weisbrod, B.A. 1961. The valuation of human capital. *Journal of Political Economy,* **69** (5): 425-436

West D: Spending pressures spark call for NICE restraints. *Health Service Journal*; 2009.

Whynes, D.K., Wolstenholme, J.L., Frew, E. 2004. Evidence of range bias in contingent valuation payment scales. *Health Economics,* **13(2)**: 183-90

Willig, R. 1976. Consumer's surplus without apology. *American Economic Review,* **66:** 589-97

Yeung, R., Smith, R.D., McGhee, S.M., 2003. Willingness to pay and size of health benefit: an integrated model to test for 'sensitivity to scale'. *Health Economics Letters* **12**, 791–796.

Zethraeus, N. 1998. Willingness to pay for hormone replacement therapy. *Health Economics,* **7** 31-38.

Zhao, F., Yue, M., Yang, H., Wang, T., Wu, J., Li, S. 2011. Willingness to Pay per Quality Adjusted Life Year: is One Threshold Enough for Decision making? Results from a study with chronic prostatitis. *Medical Care,* **49:** 267-272.

Appendix A2 – Descriptions of existing MVQ studies (not including VPF based studies which are adequately addressed in the text)

A2.1 Studies eliciting values from the general population

Johannesson and Johansson (1997a) ask 2824 members of the Swedish general population between the ages of 18 and 69 their WTP for an increase in their life expectancy of 1 year. Interviews were conducted on the telephone. Respondents were told the chance they would survive to the age of 75 (varied depending on the age and sex of the respondent) and then told that upon reaching 75 they could have a treatment that would increase their life expectancy from 10 years to 11 years (i.e. live to 86 rather than 85). Respondents were asked if they would pay a given insurance premium immediately (respondents were randomised to 1 of 6 different values). In an attempt to try and control for people overestimating their WTP value as a result of the hypothetical nature of the exercise (Diamond and Huasman, 1994), respondents were asked a follow up question to confirm that they were absolutely sure that they would (not) pay. To test whether the observed WTP value is influenced by the expected quality of life in later life respondents were asked how high they thought their quality of life would be, on a scale from 1 to 10, during an extra life year at an advanced age.

Through ordered logit regression analysis Johannesson and Johansson (1997a) estimated the maximum insurance premium the average respondent was willing to pay was about 9300 Swedish Kroner (£1,100) according to the standard yes/no question. According to the follow up question the value was 4700 SEK (£600). No value of a QALY is presented but given that the mean anticipated QoL in the additional year was 0.45 (on the QALY scale), a crude estimate would be **£2,000** according to the standard question and **£1,000** according to the follow up question. (Note these values differ to those presented in Mason *et al.* 2008 as they do not divide the value by the QoL associated with the extra year of life). They also found a strong relationship between expected quality of life in the additional year of life and the WTP value obtained. They argue this is evidence of strong sensitivity to scope of the WTP instrument. However, this quality of life consideration is not the valuand explicitly mentioned in the WTP question, but only an implicit judgement made by the respondent. A better test of scope would have been to vary the number of years added to life expectancy, both between respondents and within respondents. The results suggest there may have been some bias introduced by the use of the discrete choice WTP approach. Considering responses to the more rigorous follow up question, 68% of respondents faced with the smallest premium would not pay. This suggests that the study is not adequately capturing respondent's WTP at the lower end. At the upper end only 4% of those presented with the highest value accepted it, suggesting that the upper end is more adequately covered.

Johannesson and Johansson (1997a) do not adjust their results to allow for time preference. This is surprising, especially given that they observe a positive relationship between age and WTP, which they explain by a positive rate of time preference estimated to be about 1%. This omission suggests their estimates are lower than the true values. Another explanation offered for the very low estimates, is that since many of the respondents have children (actual proportion is not presented) immediate risk reductions, as valued in the VPF estimates, have an 'extra' value over and above the value derived from risk reductions at an advanced age. Furthermore, studies have shown a strong public preference for devoting resources towards saving young lives rather than old lives (Cropper *et al.,* 1994; Johannesson and Johansson, 1997b).

Johnson *et al.* (1998) also estimate the value of a year of additional life at the end of an individual's expected lifetime using a sample of 246 members of the Canadian general population. They describe the method they use as WTP, but it is in fact a form of conjoint analysis. Rather than state a WTP value, respondents must chose from two options, which each have four attributes: symptoms, longevity, cost over the next three years and daily activities. A series of pair wise choices were presented to respondents and they were asked to state whether option A or option B was preferred on a 7 point scale, ranging from 'A is much better' to 'B is much better'. WTP values are estimated using ordered probit models.

Johnson *et al.* (1998) found the WTP for a year of additional life with minimal activity restrictions to be about **£8,000**. No value for 'minimal activity restrictions' is presented making it impossible to infer a QALY value. If activity restrictions were higher, WTP for an additional year approached zero. As with the study by Johannesson and Johansson (1997a), Johnson *et al.* (1998) acknowledge the likely importance of time preference but do not apply any discounting to the results. Further explanations for the low estimates are that tradeoffs were elicited relative to marginal increases in longevity rather than for a substantial part of a person's adult life. The authors also believe that asking respondents to pay now for a benefit a large number of years in the future may reduce overall WTP as respondents may have believed the benefit to be uncertain.

Gyrd-Hansen (2003) elicits the WTP for health improvements, as measured by the EQ-5D descriptive system, from 3201 members of the Danish population. The study uses the same 42 health states employed in the derivation of the UK EQ-5D population value set (Dolan, 1997). Respondents were first given a discrete choice between two health states (the decrement between the two health states ranged from 0.012 to 0.312 on the QALY scale, as indicated by the UK value set). They were then asked if they would pay a specified monthly out of pocket payment (to which they were randomly allocated) to have a medicine that would improve their health from the poor EQ-5D state to the better one. Discrete choice modelling was used to analyse the data. Two WTP per QALY values are presented. The first, based on the Danish EQ-5D value set is **£8,000**. The second, based on the UK EQ-5D value set is **£7,000**.

This study could have been influenced by the methodological issues surrounding WTP. Of respondents presented with the highest amount (10,000DKK=£1,350) 15.9% were willing to pay more than this amount for the health improvement, which suggests that the choice of values was not appropriate to capture the maximum WTP. Furthermore, 17.9% of respondents presented with the smallest amount (100DKK=£13.50) refused to pay it. The author thinks it unlikely that these are protest responses, but instead feels that lower prices should have been used to more accurately capture the lower end of the WTP spectrum. The author also suggests that the results may have been insensitive to scope. She found that in four out of the five dimensions moderate problems did not influence preferences as indicated by WTP. The author states that, 'these results suggest that the ability of WTP to discriminate between differences in health status is less than for TTO and discrete choice experiments' (p.1057). The author also acknowledges the potential problems of hypothetical bias and the constraint imposed by ability to pay. She finds the effect of ability to pay to be minor and consider uncovering any hypothetical bias to be beyond the scope of the study.

The QALY value found by Gyrd-Hansen (2003) is lower than that found by studies based on preferences for reducing the risk of death (VPF studies). The main explanation offered for this is that valuations of reducing risks of death operate with lives saved as the outcomes, rather than quality of life improvement as in the Gyrd-Hansen study. Although the TTO method elicits trade-offs between quality and length of life, Gyrd-Hansen argues that the apparent inconsistency between her study and risk of death studies suggests that the relative weighting of quality and length of life may differ. She also acknowledges the dubious nature of extrapolating the WTP values for small increments in health to scenarios in which greater health improvements, or life extensions, are provided.

The results in Gyrd-Hansen (2003) find that WTP per QALY decreases with age (confirming a similar finding by Olsen and Donaldson, 1998). One possible explanation offered is that with increasing age there is a greater propensity to accept deteriorating health. Alternatively, older people may be more hesitant to consume medical treatment. No discount rate is applied in this study. The author argues that since payment and health improvements occur simultaneously (through ongoing monthly payments) WTP will only be affected by choice of time period in so far that time preferences for health differ markedly from time preferences for other goods. Published economic evaluation guidelines now often recommend that costs and health benefits be discounted equally (NICE, 2008).

Byrne *et al.* (2005) presented 193 members of the US public with two osteoarthritis scenarios, descriptions of which were based on the EQ-5D domains. Utility values for the health states were elicited through VAS, TTO and SG. WTP to move from each of the two health states to full health, as well as from own health to full health, was elicited through an open-ended lump sum question. The WTP per QALY varied substantially depending on the utility measurement technique used. Also, movement from own health to full health elicited higher values than either of the hypothetical values. Across all methods, applying a discount rate of 3% the range was from £900 to £4,000 per QALY. The lowest value calculated was for the most severe health state using SG, and the highest value calculated was for own health using TTO.

Higher income categories were often associated with higher WTP/QALY values. The main explanation the authors offer for their results being lower than most previous studies is that their study involves no risk of death while most previous studies do. They feel people tend to overvalue small risks of death. The authors acknowledge a number of weaknesses of the WTP approach used in their study. Most notably, since a one-off lump sum was used as the payment method, ability to pay may have been more of a constraint than if payment had been facilitated on a recurrent basis. Surprisingly, the authors suggest that using an open-ended question may have biased the results downwards. As will be seen in the following section, the majority of evidence seems to suggest that open-ended questions result in artificially high values. Finally, as a sensitivity analysis the study uses discount rates of 5% and 1%. The results show that the effect of varying the discount rate is less than £1,000.

Mason and Donaldson (2007) note that previous study results differ depending on whether respondents value QoL improvements, life extensions or life saving treatments (or risk of death VPF studies). To clarify, a life extending treatment is one that adds a specified amount of time onto the end of an individual's life expectancy, while a life saving treatment is one that prevents imminent death. The authors test the hypothesis that WTP per QALY gained will vary according to how the health gain arises. They tested this hypothesis by asking five different WTP questions through focus groups with members of the UK general population. The first of these (*quality of life enhancing*) asked people how much they would pay (open-ended) to avoid living in an impaired EQ-5D state for one year (11211, utility decrement of 0.117, as informed by the UK EQ-5D value set, Dolan, 1997). The second question (*life extending*) asked how much respondents would pay to gain a year of life, the gain being induced by taking one tablet at age 60. The wording of the question is rather complex. The authors wanted to avoid respondents interpreting this year as a year at the end of their life as they may assume it would be in poor health and under-value it. Respondents were therefore told, "It is as if you spent (lived) a little bit extra at each age due to a slowing down of the ageing process. So at each age you experience this time in normal health for that age. These gains cause life expectancy to increase by 1 year". This wording seems both infeasible and very difficult to imagine for respondents. The third question (*life extending)* aimed to be more relevant to younger members of the sample. Therefore, respondents were asked their WTP to avoid falling into a coma in the next year of their life. Respondents were told that after the coma they would return to full health. They were also told that prior to the coma they would have time to put their affairs in order. Once again, this question seems over-elaborate.

The fourth question (*life saving)* tells respondents to imagine they have been diagnosed with a terminal illness which they will die from within a few months if they do not receive treatment. They are then asked their WTP for a treatment that would give them an extra 4 months of life with quality of life at 0.25 (as they had already been familiarised with the QALY, the health state could be described to them in this manner). Finally, the fifth question (*life saving)* is an elaborate risk reduction question. Respondents are told they have sustained injuries and if the emergency services arrive within 20 minutes they will have a 1% chance of death. They are then asked their WTP to have the emergency services arrive within 10 minutes which would reduce the chance of death to 0%.

There were a total of 52 respondents who answered all five questions. Payment cards (15 of them) were used and payment was through an immediate lump sum. The study seems too complex, especially considering that each focus group began with an explanation of finite resources in the NHS, the nature of the QALY as a benefit measure and a Q-sorting exercise. Indeed, the authors acknowledge that this took one hour to complete which left only 30 minutes to complete the WTP exercise. Furthermore, each WTP question was rather more complex than is outlined above, with each question mentioning the possibility of taking out a loan to make payment (to reduce the limit imposed by ability to pay presumably). This may partly explain the large number of zero responses. When such a response was given respondents were asked to write a short statement explaining their reasoning. From this information the authors were able to determine that the majority of zero responses were not protest responses, but indicated that respondents really did not value the treatment presented in the scenarios. For the quality of life improving question there were 22 zero responses of which 4 were protest responses. For the first life extending question (1 Equivalent Years Gain) there were 21 zero responses (2 protest responses), for the second life extending questions (avoid 1 Compensating Years Loss) there were 15 zero responses (3 protests). For the first life saving question (extra 4 months of life in poor health) there were 32 zero responses (4 protests), and in the second life saving questions (eliminate 1% risk of death) there were 11 zero responses (6 protests). Excluding protest responses the mean WTP per QALY from the quality of life improving question was **£5,000**. The WTP per QALY value for achieving a 1 Equivalent Years Gain in life expectancy was **£2,000**, while the value for avoiding a Compensating Years Loss in life expectancy was **£6,000**. The WTP per QALY value in the two life saving questions were **£19,000** and **£25,000**.

The results broadly confirm Mason and Donaldson's hypothesis that WTP per QALY would be higher for life saving treatments than life extending treatments or quality of life improving treatments. Furthermore, the results from this study are much lower than previous work by the authors based on the VPF (Mason *et al.* 2005), which suggests people value health interventions differently to safety interventions. The authors argue a different CE threshold is required for life saving treatments than is used for life extending or quality of life enhancing treatments. Furthermore, differences in values between the two life extending questions, and between the two life saving questions, is evidence of possible framing effects.

No allowance for time preference is made in this study. This would only affect question 2 (gain in life expectancy), and this is perhaps unsurprisingly the question that elicits the lowest MVQ estimate. Allowing for time preference would increase this value and make it more similar to those estimated through the other questions.

Donaldson *et al.* (2008) published a report on the UK based Social Value of a QALY project. As well as attempting to estimate equity based QALY weights, this project seeks to estimate an MVQ. The study uses two health states, one involving recurrent stomach bowel problems, the other involving recurrent episodes of head pain. Three different durations are used for each health state: 3 months, 12 months and 'the rest of your life'. Respondents were asked their WTP to either prevent the certainty of the given illness or to eliminate some risk of the illness (either 10% or 5%). Following the WTP part of the questionnaire, values for the health states were obtained through SG. In the WTP exercise, respondents were given a pack of 16 payment cards ranging from £1 to £1 million and asked to sort them into three piles: those amounts they definitely would pay, those they definitely would *not* pay, and those about which they were unsure. Respondents who would not pay anything were asked to explain their reasons.

The WTP results are relatively insensitive to scope. Mean WTP to avoid 12 months with stomach pain is only 2.3 times greater than WTP for 3 months of stomach pain. Mean WTP to avoid 12 months with headaches is only 2.2 times greater than WTP to avoid 3 months with headaches. Considering the risk based questions, there is even less sensitivity to scope. In the case of headaches the WTP to avoid 5% chance of having the condition was higher than the WTP to avoid a 10% chance of developing the condition. This finding held for risk of 3 months with the illness, 12 months and lifetime.

Donaldson *et al.*(2008) raise an interesting point regarding the method of calculating the MVQ, which is relevant to the calculation method used in the following chapter. When a study elicits both a WTP for a health improvement and directly elicits a utility value for the given health improvement, how should the information be combined to generate an MVQ estimate. The approach more closely aligned with welfare economic theory, since it allows for individual variation in results, would be to estimate an MVQ for each individual and then calculate an average of the individual estimates (the 'individual approach'). However, in the Donaldson study since some respondents are only prepared to accept a very small amount of risk in the SG exercise they generate extremely high MVQ estimates. A total of 115 respondents (out of 403) give answers that generate estimates of more than £1 million, and some of the estimates are thousands of millions of pounds. This leads to mean MVQ estimates of £300 million for the sample valuing stomach problems, and £700 million for those valuing headaches (12 months). Therefore, Donaldson *et al.* (2008) propose an alternative which is to calculate the mean WTP, and the mean QALY gain and combine the two (the 'aggregate approach'). This approach is similar to the approach used in the VPF studies, which make no attempt to estimate individual based MVQ's.

If one wishes to subscribe to the conventional precepts of welfare economics and if one could be confident that individuals have values and preferences that conform with standard assumptions, and that their responses reveal those values with total accuracy and precision, the 'individual approach' would be appropriate. However, if responses contain elements of bias, these will be magnified under the 'individual approach' and there may be a case for employing the 'aggregate approach'. Clearly this distinction becomes more important when very small trades are permitted in the preference elicitation exercise. The problem could be limited by increasing the smallest unit of trade allowed in the exercise. One must be cautious when making such a move though, as it can result in an increased number of problematic 'non-trades'.

Of the previously mentioned studies, the distinction between the 'individual approach' and the 'aggregate approach' is only potentially relevant for the studies by Johannesson and Johansson (1997) and Byrne *et al.* (2005). In the Johannesson and Johansson (1997) study no MVQ is presented so one was estimated through the 'aggregate approach'. In the Byrne *et al.* (2005) study, the 'individual approach' is used. However, the highest mean utility value was 0.858 which perhaps suggests that the effect of very small utility decrements was of limited importance.

No discount rate is applied in the Donaldson *et al.* (2008) study. Once again, this is likely to lead to over-estimates for the questions involving conditions that persist for the rest of your life. The median MVQ estimates for the 'individual approach' are £26,000 based on 12 months of stomach problems, and £57,000 based on 12 months of headaches. Using the aggregate approach the mean MVQs are £17,980 based on 12 months of stomach pain, and £22,570 based on 12 months of headaches. Across all different combinations of illness type, standard gamble questions, and WTP questions and using both aggregation procedures and means and medians, the MVQ estimates range from **£18,000** to **£700 million**.

Prades *et al.* (2009) use a complex study design that allows them to test the linearity of their MVQ estimates with respect to severity of different conditions, duration of a given condition and reductions in the risk of chronic conditions. A total of 13 different questions were administered through interviews with 892 members of the Spanish general public. In 9 of these questions respondents were asked to assume that they had an illness that would put them in a specified impaired EQ-5D health state for the remainder of their lives (either state 22223 or state 21212). The two EQ-5D states were valued by the respondents through SG earlier in the exercise. They were also told they could have a treatment (A) for free that would cure the illness, but it would take either 2 weeks, 2 months or 4 months to take effect. They were then asked their WTP for a better treatment (B) which would either work more quickly, or reduce the severity of the symptoms experienced prior to the cure. The other 4 questions involved the WTP for a drug that would reduce the chance of developing an illness, from 1% to 0% or from 1% to 0.5%. The WTP format was payment card. Respondents were presented with 11 different monetary monthly payments, that would last for a period of 12 months, in a random order and asked if they would pay them. Respondents who were prepared to pay the highest amount were asked how much more they would pay. To test the effect of payment duration one group of respondents were told the payment period would last for 24 months instead of 12 months.

The method of calculating an MVQ is the 'aggregate approach'. Prades et al. (2009) find alarming variation in MVQ values by factors such as size of the health gain, period of payment and duration of the health impairment. For the first 9 questions the MVQ ranges from **£8,000** to **£110,000**. For the risk based questions the estimates range from **£4,000** to **£12,000**. In light of these results Prades and colleagues offer quite a damning assessment of the potential for producing a robust all-purpose MVQ figure. The reasons they give for this are insufficient sensitivity to the duration of health states and the size of QoL improvements and substantial order effects and payment period effects. Order effects arise when the value obtained is influenced by the order in which the questions are presented (tested through between respondent comparisons). The explanation they offer for lack of sensitivity is potential budget constraints, but given that the health gains and durations used are rather modest this explanation (if valid) would pose serious questions over past and future WTP for a QALY studies.

No allowance for time preference is made in the study by Prades *et al.* (2009). This is not a problem in the 9 questions that reduce the time spent in EQ-5D health states by a maximum of 4 months, and in which the payment period is 12 months. However, in the risk based questions the respondents could potentially remain in the impaired health state for the rest of their lives in which case some discounting is appropriate. Without discounting the estimates from these particular questions may be under-estimates.

A2.2 Studies eliciting values from patients

Zetheraeus (1998) asked 104 Swedish women between the ages of 45 and 65, who were receiving hormone replacement therapy (HRT), their WTP, through monthly out of pocket payments, to continue receiving HRT. Quality of Life without HRT was assessed through both VAS and TTO. The WTP questions were binary closed-ended questions i.e. respondents were presented with one monetary amount and asked if they would pay it. Eight different monetary amounts were used ranging from SEK 100 to SEK 10,000. Mean WTP values were estimated through both parametric and non-parametric methods. Two logistic regressions were used to estimate the WTP parametrically, the first based on the VAS responses, the second based on the TTO responses. Non-parametrically the proportion of yes responses at each price level are used to construct a curve that shows the relationship between the price and the proportion of yes answers. This curve is interpreted as a demand curve and the mean WTP is measured as the area below the curve.

The TTO based logistic regression estimates the monthly WTP to be £300. Combining this with a utility decrement of 0.29, using the 'aggregate approach', gives an MVQ estimate of **£12,000**. The VAS based logistic regression estimates the monthly WTP to be £280. Combining this with a utility decrement of 0.37 gives an MVQ estimate of **£9,000**. Finally, the non-parametric approach estimates the mean WTP to be £270. This generates MVQ estimates of **£12,000** and **£9,000**, using the TTO and VAS based utility decrements respectively. No discount rate is applied in this study. Since both the costs and benefits occur simultaneously and the time horizon is relatively short, this should not bias the results.

Blumenschein and Johannesson (1998) asked 69 US asthma sufferers to value their own current health using SG, VAS and TTO. They were then asked their WTP for an improvement in health from current health to perfect health. Two WTP elicitation formats were used: closed-ended dichotomous choice and a bidding game. The dichotomous choice question was always asked first, followed by the biding game. The values used in the bidding game are not presented. The dichotomous choice results are analysed using logistic regression to estimate a maximum WTP. The mean health state utilities are 0.68, 0.89 and 0.91 for the VAS, TTO and SG respectively. The mean monthly WTP for a cure is £140 through the bidding game, and £260 through the dichotomous choice questions. Combining the utilities with the WTP, through the 'aggregate approach' gives MVQ estimates ranging from **£5,000** to **£35,000** depending on the method used to estimate WTP and utilities. No discounting is used as both the costs and benefits are incurred at the same time.

The study by Blumenschein and Johannesson (1998) highlights the effect of the WTP elicitation method used (further evidence on which will be presented in the following section). The results lead to an MVQ estimate based on TTO derived utilities and the dichotomous choice questions of £28,000. However, basing the MVQ estimate on TTO derived utilities and the bidding game questions leads to a figure of £16,000.

Cunningham and Hunt (2000) ask 40 patients in the UK with dentofacial deformities to value their health state through SG. They were then asked their WTP to undergo treatment to correct their deformity. Payment cards were used with values ranging from £0 to £15,000 in £1,000 increments. Payment was out of pocket, and respondents were told they could pay in instalments if they wished. The mean WTP is £6,833. The mean health state utility is 0.73. A QALY value is calculated by assuming a remaining life expectancy of 50 years. No information is given on how this value was obtained. Given the mean age of the sample is 24, 50 years of remaining life expectancy is likely to be an underestimate. The MVQ estimate (generated through the 'aggregate approach') is **£600**. This does not allow for any discounting and so is likely to be an underestimate (although this may be cancelled out by the underestimate of remaining life expectancy).

King *et al.* (2005) asked three US patient populations to value their own current health using SG, TTO and VAS. The three populations consisted of general medical patients (n=117), patients with a degenerative spine condition (n=84) and patients with cerebral aneurysms (n=191). They were then asked their WTP for an improvement in health from current health to perfect health. An iterative closed-ended bidding method was used. Payment was through a one-off immediate lump sum payment. The values presented to respondents in the bidding game ranged from $1 to a value of 10 times the subjects own household income (a computer programme was used to make these calculations). Cumulative lifetime QALYs being purchased were calculated by multiplying the difference between current health and perfect health by life expectancy, as garnered from US life tables. A discount rate of 3% was applied to future health benefits, while the one-off WTP format meant no discounting of costs was necessary. The WTP values were then combined with the QALY gains to give WTP/QALY ratios (using the 'individual approach'). They estimate MVQs ranging from £9,000 to £23,000. MVQ estimates were typically lower when based on VAS valuations, while estimates based on SG and TTO valuations were very similar.

The role of respondent income is particularly important in the study by King *et al.* (2005) since the WTP questions are framed in relation to own household income. Indeed, regression models predicting WTP/QALY calculated with the VAS, SG, and TTO showed patient income was consistently associated with WTP/QALY ratios – wealthier patients had higher ratios. In the SG model an income change from $25,000 to $50,000 would increase the predicted WTP/QALY ratio by $64,100. Since the study population had median incomes below that of the US population the authors acknowledge that their MVQ estimates may be under-estimates.

It is potentially possible to use past WTP studies to derive an MVQ retrospectively. For this to be possible the study must either elicit the WTP for the alleviation of a condition for which a utility value already exists, or also elicit utility values. **Lundberg *et al.* (1999)** is an example of one such study. They ask 366 Swedish patients with either psoriasis or atopic eczema to value their current health state using VAS, TTO and SG. They then ask the respondents their WTP for a cure of their condition, using both dichotomous choice and bidding game formats. Payment was through monthly out of pocket. They found that the dichotomous choice questions yielded the highest WTP values. Although not explicitly presenting an MVQ figure, the utilities and WTP values can be combined (through the 'aggregate approach') to generate MVQs (done by CT). The lowest value is generated by combining the VAS derived utility for psoriasis (0.69) with the bidding game derived monthly WTP for a cure of this condition, giving a value of £5,000. The highest value is generated by combining the SG derived utility for psoriasis (0.97) with the dichotomous choice derived monthly WTP for a cure of this condition, giving a value of **£88,000**.

Table A4.1 - Background characteristics by whether or not respondents gave LG responses in the Income Loss questions

		Traders	LG responders	Chi² Test (p-values)
Number of Respondents		105	136	
Gender	Male	53%	51%	0.774
	Female	47%	49%	
Age	Average	43.74	42.76	
	SD	12.99	13.37	
	18-35	29%	35%	0.568
	36-50	34%	29%	
	51-65	37%	36%	
Educated beyond the minimum school leaving age	Yes	60%	71%	**0.085**
	No	40%	29%	
Educated to Degree Level	Yes	29%	35%	0.323
	No	71%	65%	
Employment	Employed	48%	58%	**0.026**[1]
	Self-Employed	3%	6%	
	House Wife/Husband	17%	9%	
	Pensioner	9%	6%	
	Work Seeking	3%	3%	
	Unable to Work	11%	9%	
	Student	9%	9%	
Net Own Monthly Income	<1000 Euros	40%	37%	0.932
	1000 - 1499	22%	21%	
	1500 - 1999	17%	20%	
	>2000 Euros	21%	22%	
Children	Yes	49.0%	50%	0.826
	No	51.0%	50%	
Religion	Protestant	16%	17%	0.573
	Roman Catholic	26%	31%	
	Atheist	53%	46%	
	Other	5%	6%	
Marital Status	Married	45%	41%	0.577[2]
	Single/Never Married	20%	24%	
	Divorced	12%	12%	
	Widowed	3%	2%	
	Living Together	19%	17%	
	Other	1%	4%	
Mean Self-Reported Health on the EQ-VAS[3]		0.71	0.77	**0.058**

1. Employment treated as dichotomous, employed/self-employed vs the rest.
2. Marital status treated as dichotomous, married vs the rest.
3. Sample Size: Traders (94), LG responders (119)

Table A4.2 - Background characteristics by whether or not respondents traded in the Income Gain questions

		Traders	LG responders	Chi² Test (p-values)
Number of Respondents		87	154	
Gender	Male	49%	54%	0.505
	Female	51%	46%	
Age	Average	42.87	43.36	
	SD	12.91	13.38	
	18-35	32%	32%	0.863
	36-50	33%	30%	
	51-65	35%	38%	
Educated beyond the minimum school leaving age	Yes	64%	67%	0.692
	No	36%	33%	
Educated to Degree Level	Yes	30%	33%	0.605
	No	70%	67%	
Employment	Employed	53%	56%	0.288[1]
	Self-Employed	6%	5%	
	House Wife/Husband	13%	10%	
	Pensioner	6%	6%	
	Work Seeking	3%	3%	
	Unable to Work	11%	10%	
	Student	8%	10%	
Net Own Monthly Income	<1000 Euros	41%	36%	0.657
	1000 - 1499	17%	24%	
	1500 - 1999	20%	18%	
	>2000 Euros	22%	22%	
Children	Yes	44%	53%	0.183
	No	56%	47%	
Religion	Protestant	18%	16%	0.704
	Roman Catholic	24%	31%	
	Atheist	52%	48%	
	Other	6%	5%	
Marital Status	Married	43%	43%	0.961[2]
	Single/Never Married	25%	21%	
	Divorced	8%	14%	
	Widowed	1%	3%	
	Living Together	22%	15%	
	Other	1%	4%	
Mean Self-Reported Health on the EQ-VAS[3]		0.71	0.76	0.161

1. Employment treated as dichotomous, employed/self-employed vs the rest.
2. Marital status treated as dichotomous, married vs the rest.
3. Sample Size: Traders (76), LG responders (137)

Table A4.3 - Background characteristics by whether or not respondents over-traded in the Income Loss questions				
		Over-Traders	Rest of the sample	Chi2 Test (p-values)
Number of Respondents		40	201	
Gender	Male	52%	52%	0.976
	Female	48%	48%	
Age	Average	43.95	43.03	
	SD	11.03	13.60	
	18-35	25%	33%	0.528
	36-50	38%	31%	
	51-65	37%	36%	
Educated beyond the minimum school leaving age	Yes	60%	67%	0.382
	No	40%	33%	
Educated to Degree Level	Yes	22%	34%	0.160
	No	78%	66%	
Employment	Employed	55%	53%	0.888[1]
	Self-Employed	2%	5%	
	House Wife/Husband	15%	12%	
	Pensioner	3%	8%	
	Work Seeking	0%	3%	
	Unable to Work	22%	8%	
	Student	3%	11%	
Net Own Monthly Income	<1000 Euros	35%	39%	0.803
	1000 - 1499	25%	21%	
	1500 - 1999	15%	19%	
	>2000 Euros	25%	21%	
Children	Yes	40%	51%	0.194
	No	60%	49%	
Religion	Protestant	20%	16%	0.748
	Roman Catholic	30%	28%	
	Atheist	8%	5%	
	Other	42%	51%	
Marital Status	Married	40%	43%	0.701[2]
	Single/Never Married	23%	22%	
	Divorced	20%	10%	
	Widowed	5%	2%	
	Living Together	12%	19%	
	Other	0%	4%	
Mean Self-Reported Health on the EQ-VAS[3]		0.68	0.76	0.144[4]

1. Employment treated as dichotomous, employed/self-employed vs the rest.
2. Marital status treated as dichotomous, married vs the rest.
3. Sample Size: Over-Traders (33), Rest of the Sample (180)
4.This is a t-test

Table A4.4 - Background characteristics by whether or not respondents over-traded in the Income Gain questions

		Over-Traders	Rest of the sample	Chi2 Test (p-values)
Number of Respondents		45	196	
Gender	Male	49%	53%	0.613
	Female	51%	47%	
Age	Average	41.87	43.49	
	SD	12.55	13.34	
	18-35	31%	32%	0.566
	36-50	38%	30%	
	51-65	31%	38%	
Educated beyond the minimum school leaving age	Yes	69%	65%	0.647
	No	31%	35%	
Educated to Degree Level	Yes	33%	32%	0.825
	No	67%	68%	
Employment	Employed	60%	52%	0.575[1]
	Self-Employed	2%	5%	
	House Wife/Husband	9%	13%	
	Pensioner	5%	8%	
	Work Seeking	2%	3%	
	Unable to Work	11%	10%	
	Student	11%	9%	
Net Own Monthly Income	<1000 Euros	40%	38%	0.498
	1000 - 1499	13%	23%	
	1500 - 1999	22%	18%	
	>2000 Euros	25%	21%	
Children	Yes	40%	52%	0.163
	No	60%	48%	
Religion	Protestant	22%	15%	0.650
	Roman Catholic	24%	30%	
	Atheist	47%	50%	
	Other	7%	5%	
Marital Status	Married	36%	44%	0.280[2]
	Single/Never Married	29%	21%	
	Divorced	11%	12%	
	Widowed	0%	3%	
	Living Together	22%	17%	
	Other	2%	3%	
Mean Self-Reported Health on the EQ-VAS[3]		0.73	0.75	0.561[4]

1. Employment treated as dichotomous, employed/self-employed vs the rest.
2. Marital status treated as dichotomous, married vs the rest.
3. Sample size: Over-Traders (39), Rest of the Sample (174)
4. This is a t-test

Table A4.5 - Mean number of years traded excluding LG responders with a balanced panel

		20% (n=18)		40% (n=23)		60% (n=25)	
		Loss	Gain	Loss	Gain	Loss	Gain
Number of years traded to either avoid an income loss or achieve and an income gain *excluding* LG responders	Mean	3.09	3.38	3.93	3.74	4.16	4.06
	SD	2.97	3.42	3.09	3.21	3.37	3.54
	10th Percentile	0.25	0.25	0.25	0.25	1.00	1.00
	Median	2.00	1.67	4.00	2.42	2.00	2.00
	90th Percentile	10.00	10.00	9.00	9.00	9.92	9.92

Table A4.6 - MVQ values calculated at the individual level (excluding LG responders) based on both individual income and household income (with a balanced panel[1])

		Variant A: 20%		Variant B: 40%		Variant C: 60%	
		Loss	Gain	Loss	Gain	Loss	Gain
		(n=18)	(n=18)	(n=23)	(n=23)	(n=25)	(n=25)
Mean number of years traded		3.09	3.38	3.93	3.74	4.16	4.06
Mean Annual Individual Income (Euros)		13,717	13,717	15,605	15,605	19,998	19,998
Mean Annual Household Income (Euros)		26,272	26,272	26,083	26,083	29,517	29,517
Number of negative responses (truncated to zero)		7	9	11	11	7	10
Income equivalent to 1 Quality Adjusted Life Year (Based on Individual Income, Euros)	Mean	**21,314**	**17,306**	**67,070**	**75,286**	**60,920**	**55,290**
	SD	48,636	43,874	172,191	212,672	122,482	117,274
	10th percentile	0	0	0	0	0	0
	Median	0	1,588	0	2,039	15,984	12,000
	90th percentile	105,000	54,346	119,880	116,683	136,500	113,100
Income equivalent to 1 Quality Adjusted Life Year (Based on Household Income, Euros)	Mean	51,664	40,177	116,824	139,473	179,723	149,106
	SD	137,633	122,332	306,114	338,598	521,600	489,036
	10th percentile	0	0	0	0	0	0
	Median	0	1,800	0	4,795	15,984	14,614
	90th percentile	294,000	64,378	344,880	335,684	273,000	226,200

[1]A balanced panel means the same respondents appear in both the loss and gain questions for each level of income change

Table A4.7 - MVQ values calculated at the individual level (excluding LG responders and over-traders) based on both individual income and household income

		Variant A: 20%		Variant B: 40%		Variant C: 60%	
		Loss	**Gain**	**Loss**	**Gain**	**Loss**	**Gain**
		(n=11)	(n=12)	(n=21)	(n=14)	(n=33)	(n=16)
Mean number of years traded		1.26	0.85	1.99	1.12	2.67	1.46
Mean Annual Individual Income (Euros)		14,450	15,412	15,043	14,139	17,755	19,560
Mean Annual Household Income (Euros)		27,995	27,410	27,042	27,782	27,269	29,559
Income equivalent to 1 Quality Adjusted Life Year (Based on Individual Income, Euros)	Mean	**34,877**	**102,013**	**76,755**	**131,913**	**79,214**	**91,586**
	SD	59,168	253,057	177,911	261,924	143,303	136,432
	10th percentile	0	3,176	0	4,795	1,598	11,189
	Median	7,992	6,691	13,320	29,890	29,400	60,900
	90th percentile	105,000	182,218	119,880	372,428	279,720	224,400
Income equivalent to 1 Quality Adjusted Life Year (Based on Household Income, Euros)	Mean	84,541	155,092	137,770	240,407	210,609	238,173
	SD	170,720	322,899	315,112	408,921	522,218	599,714
	10th percentile	0	3,600	0	10,460	3,000	13,997
	Median	7,992	25,594	46,992	67,096	30,000	63,450
	90th percentile	294,000	524,218	344,880	978,601	388,493	301,517

A 5.1 Participant Information Sheet

Project Title: An exploration of the way people value life and money

You are being invited to take part in a research project. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Feel free to ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

What is the purpose of the study?

The purpose of this study is to gain a better understanding of how people value their health and money. The study is testing a new method for valuing health.

Why have I been chosen?

People are being selected at random to gain a sample of 100 members of the general population.

Do I have to take part in the study?

No. It is up to you to decide if you want to take part in the study. If you decide to take part you will be asked to sign a consent form. If you think you want to take part and then change your mind, you can do so without giving a reason.

What will be involved if I agree to take part in the study?

We will ask you to arrange an appointment to complete an interview. The interview will take about one hour. The interview will involve you answering some questions about length of life and income. You will be asked to value your own health. You will also be asked to give an indication of your salary. This information will be treated as confidential (see below).

When and where will the study take place?

A member of Oxford Outcomes will be carrying out the interview with you. The interview will be carried out at a location convenient to you e.g. your home. This will be a "one off" visit and you will not be asked to do this again.

Will the information obtained in the study be confidential?

Yes. It will not be possible to connect individual interviews to the report. The interviews will have a number and not a name to identify them for the researchers. The completed interviews will be stored in our locked cupboards at the University. You will not be able to be identified in any reports or publications. The results of this research will be used as part of a PhD thesis and may be published in an academic journal.

Has the study had ethical approval?

Yes. This project has been approved via the School of Health and Related Research's ethics review procedure. The School of Health and Related Research (ScHARR) is a department of the University of Sheffield.

What if I wish to complain about the way in which this study has been conducted?

If you have any complaints or concerns please contact me (Carl Tilling). Otherwise you can contact Dr D Fletcher, Registrar, Firth Court, Western Bank, Sheffield.

My contact details are:
Carl Tilling
Department of Economics
9, Mappin Street

Sheffield
S1 4DT
01142 223421
c.tilling@sheffield.ac.uk

You will be given copies of this information sheet and the consent form to keep.

A 5.2 Participant Consent Form

Title of Research Project: An exploration of the way people value life and money

Name of Researcher: Carl Tilling

Participant Identification Number for this project: Please initial box

1. I confirm that I have read and understand the information letter dated *[insert date]* explaining the above research project and I have had the opportunity to ask questions about the project.

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without their bearing any negative consequences. In addition, should I not wish to answer any particular question or questions, I am free to decline.

3. I understand that my responses will be kept strictly confidential. I give permission for members of the research team to have access to my anonymised responses. I understand that my name will not be linked with the research materials, and I will not be identified or identifiable in the report or reports that result from the research.

4. I agree for the data collected from me to be used in future research.

   I understand that the data will be used in a PhD thesis , may be published

   in an academic journal and may be presented at academic conferences

5. I agree to take part in the above research project.

Name of Participant	Date	Signature
Name of person taking consent	Date	Signature

(if different from lead researcher)

To be signed and dated in presence of the participant

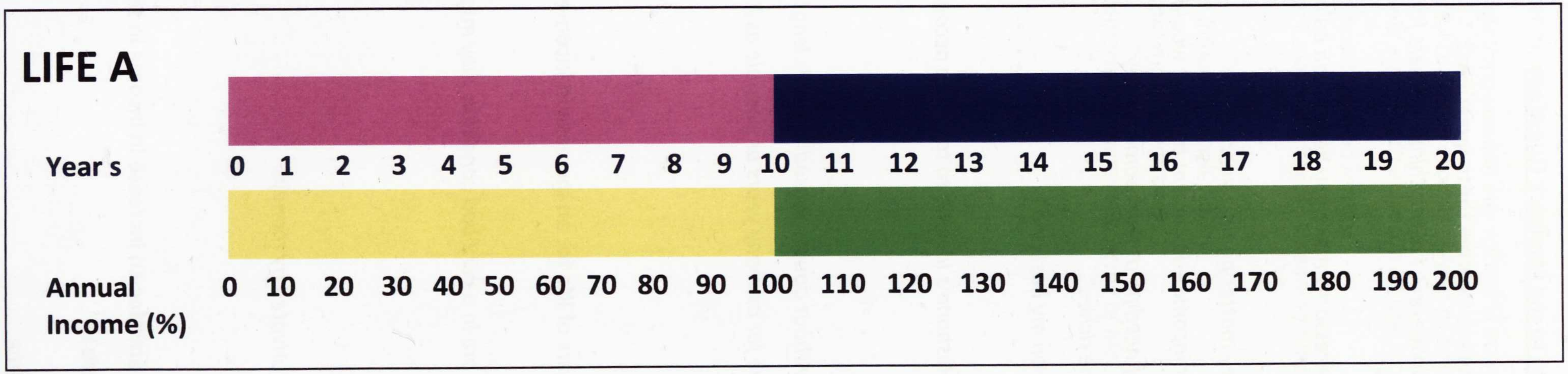
Copies:

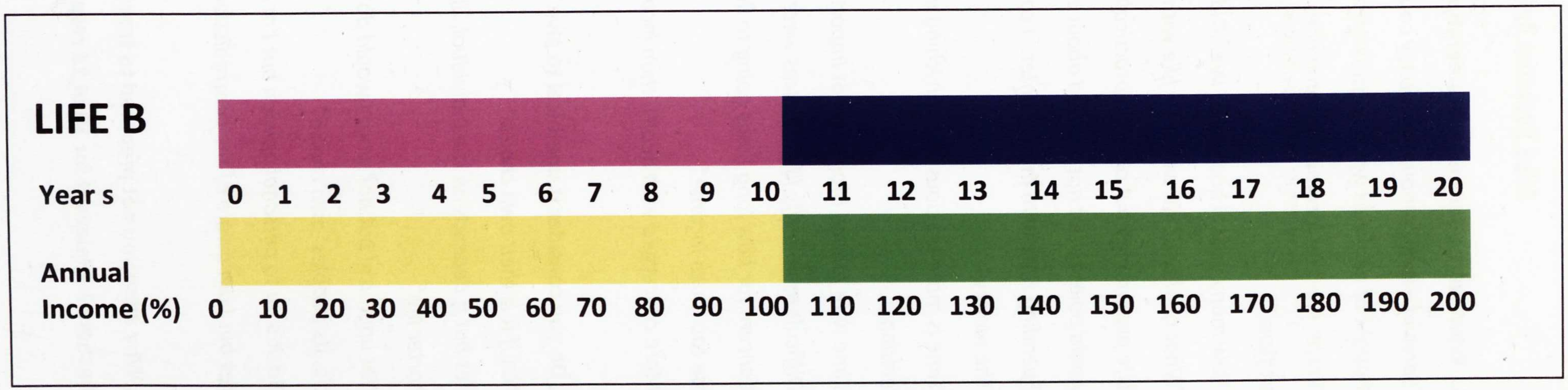
Once this has been signed by all parties the participant should receive a copy of the signed and dated participant consent form, the letter/pre-written script/information sheet and any other written information provided to the participants. A copy of the signed and dated consent form should be placed in the project's main record (e.g. a site file), which must be kept in a secure location.

A5.3 – Visual Aid

Own Current Health

Immediate Death





A 5.4 Verbatim Responses to the Feedback Questions

1. What were your main considerations when answering these questions?

"Survival – what amount of money could I survive on? I didn't think of others for money, but thought of nephews when considering time."

"Prefer life over income – income does not necessarily make you happy, can be very artificial"

"How many years I had left to live. Did money matter against years of life"

"What quality of life would be like with lower income – in current health it would be ok"

"Life over money but also the practicalities of needing at least some money"

"Avoid poverty but not worried about extreme riches"

"Spending time with my daughter. I can rely on my husband"

"The money"

"Time is more precious than anything else. Husband's income to take into account – easier decision"

"Time with family. Money is not important"

"Difficult to quantify life. Choices were not without doubt. Biased towards length of life"

"Concern for how long I was going to live. Live for as many years as possible as long as don't lose too much income"

"More concerned with health than income"

2. Do you consider it unethical to give up years of life for an increase in income?

"Not if it is their own choice"

"Yes but it depends on the individual. If they are in a very bad condition they might prefer shorter life"

"Not unethical but not what I would do"

"No, its peoples' own choice"

"No, it's up to personal opinion but I'm not motivated by money"

"Yes but I am aware of the contradiction"

3. Why were you not prepared to trade any time for an increase in income in some of the questions? (Relevant for 10 of 13 respondents)

"Value life over money"

"Life is not something you can buy – there is no substitute"

"Only live once – rather be around as long as possible"

"I have sufficient strength to enjoy life on low income"

"Life is more important than income"

"Money doesn't make you happy. Spending time with people you love gives you quality of life"

"Time is more important than money"

"Years of life are more important"

"Money not important in general scheme of life"

A6.1 VAS of Seven Combinations of Life Years and Income

As appears in the interview booklet (not seen by respondents):

"I would like to ask you to indicate how good or bad these states are. In this booklet, there is a scale like the thermometer you saw earlier.
PLACE THE CARDS ON THE ASSIGNED BOXES. DRAW A LINE FROM THE CARD '10 YEARS WITH 100% IINCOME (RA)' TO THE POINT 100 ON THE SCALE.
IF TWO CARDS HAVE BEEN RANKED AS EQUAL PLEASE MAKE A NOTE IN THE APPROPRIATE BOX ON THE VAS SCALE TO INDICATE WHICH CARD HAS BEEN PLACED IN THAT BOX.
10 years with 100% of annual income for each year is represented by 100 on the scale. The bottom end of the scale is '0 years of life, with 0% of income'.
HAND THE RESPONDENT THE SELF-COMPLETION BOOKLET OPEN AT PAGE 4.
Please indicate how good or bad each of these states is by drawing a line from the box to the scale.

CHECK THAT THE RESPONDENT DRAWS ONE LINE EACH FROM THE 6 BOXES TO THE SCALE.

IF A RESPONDENT GIVES A CARD A VALUE OF 130, ASK THEM IF THEY WOULD HAVE GIVEN A HIGHER VALUE IF THE SCALE HAD BEEN LONGER. IF THEY STATE THAT THIS IS THE CASE ASK THEM THE VALUE THEY WOULD HAVE GIVEN. PLEASE THEN WRITE THIS VALUE IN THE APPROPRIATE BOX."

Card ranked **1st** in ranking exercise

Card ranked **2nd** in ranking exercise

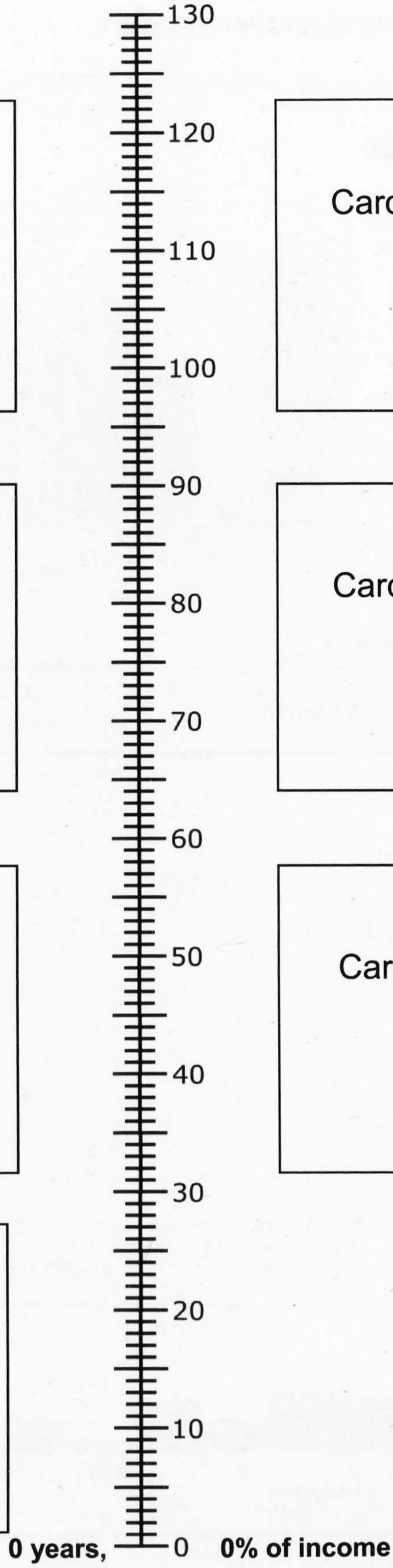
Card ranked **3rd** in ranking exercise

Card ranked **4th** in ranking exercise

Card ranked **5th** in ranking exercise

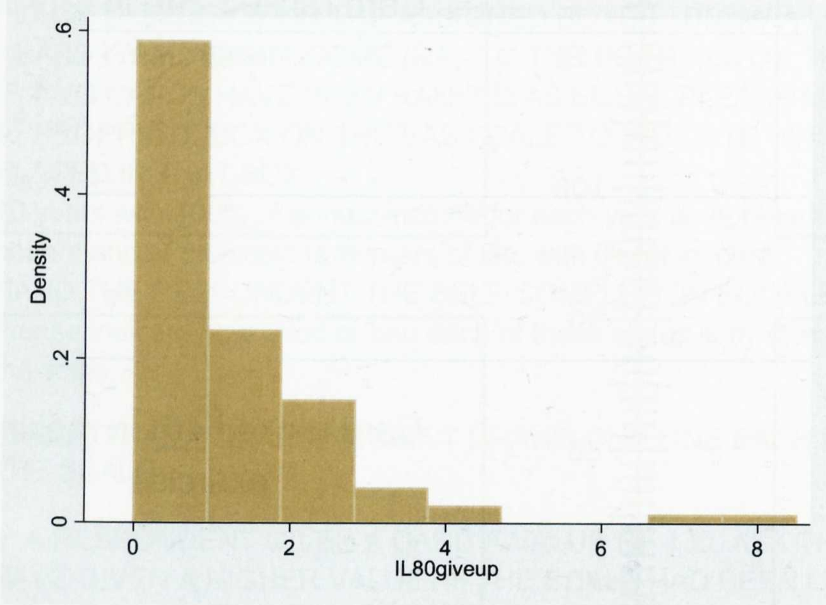
Card ranked **6th** in ranking exercise

Card ranked **7th** in ranking exercise

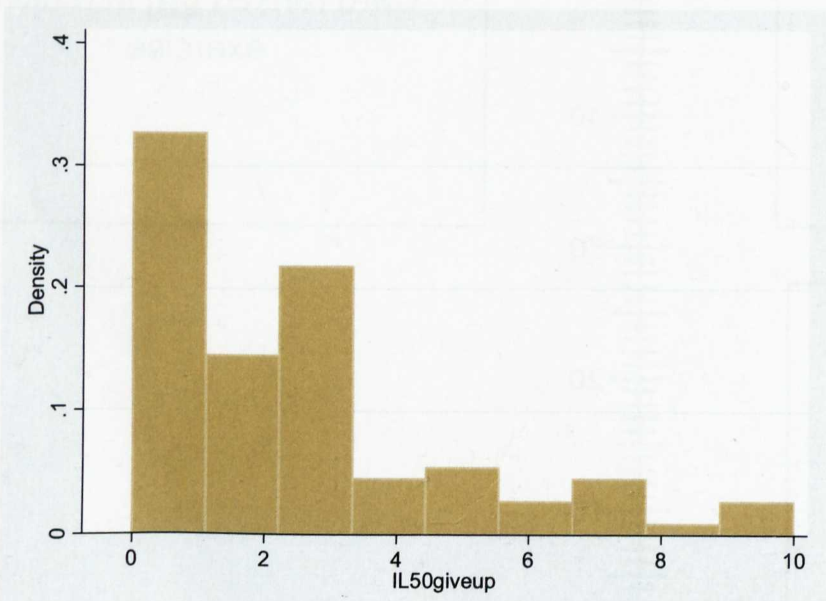


A6.2 Distribution of the Number of years/income traded

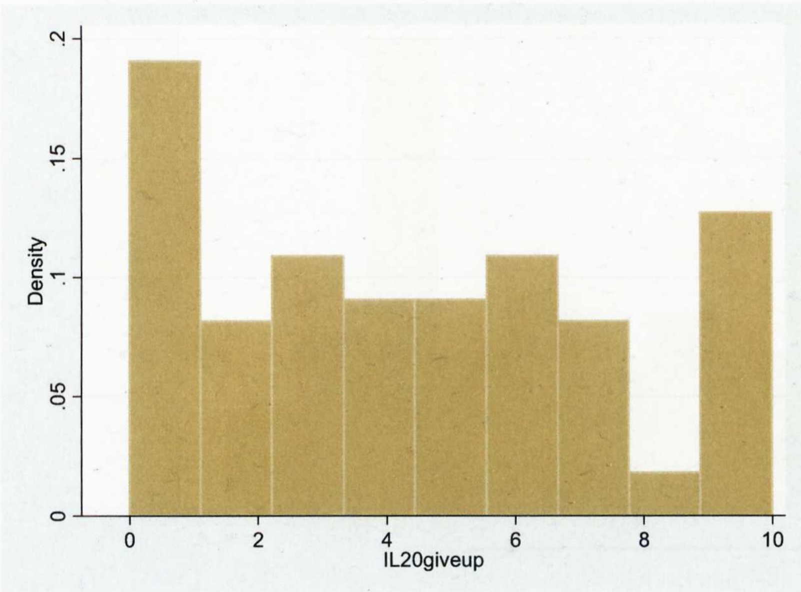
20% Income Loss (Years Traded)



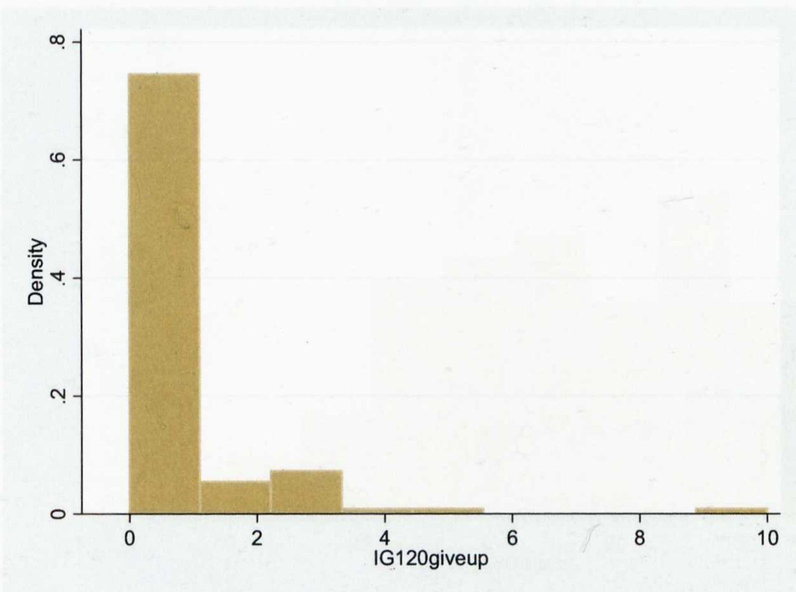
50% Income Loss (Years Traded)



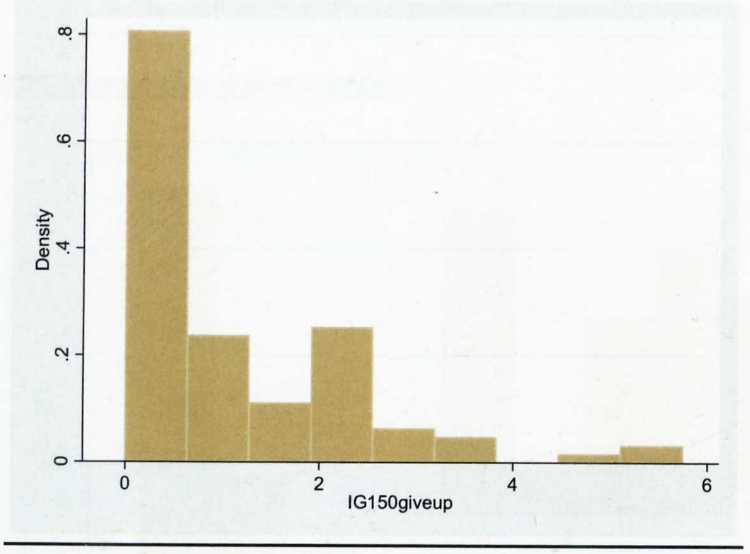
80% Income Loss (Years Traded)



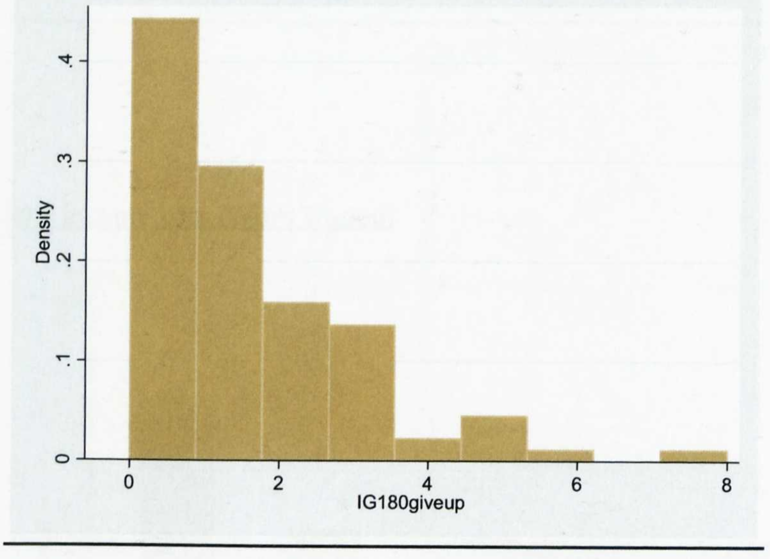
20% Income Gain (Years Traded)



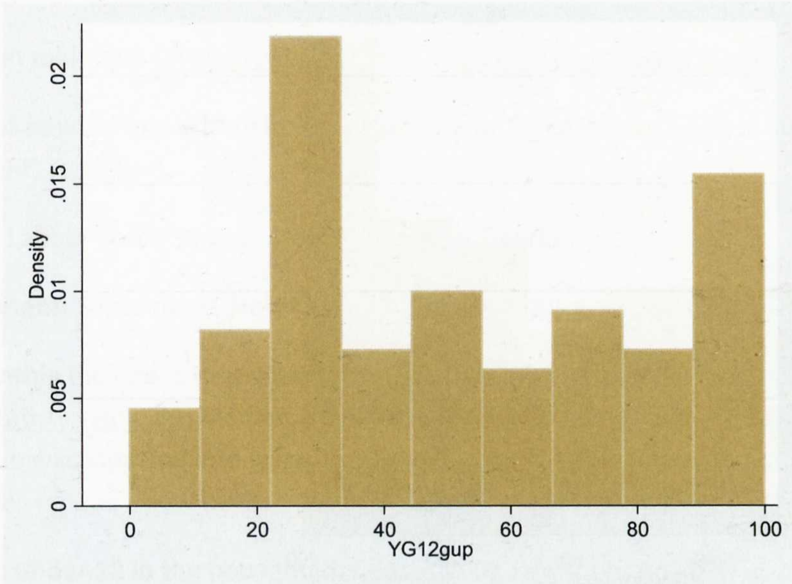
50% Income Gain (Years Traded)



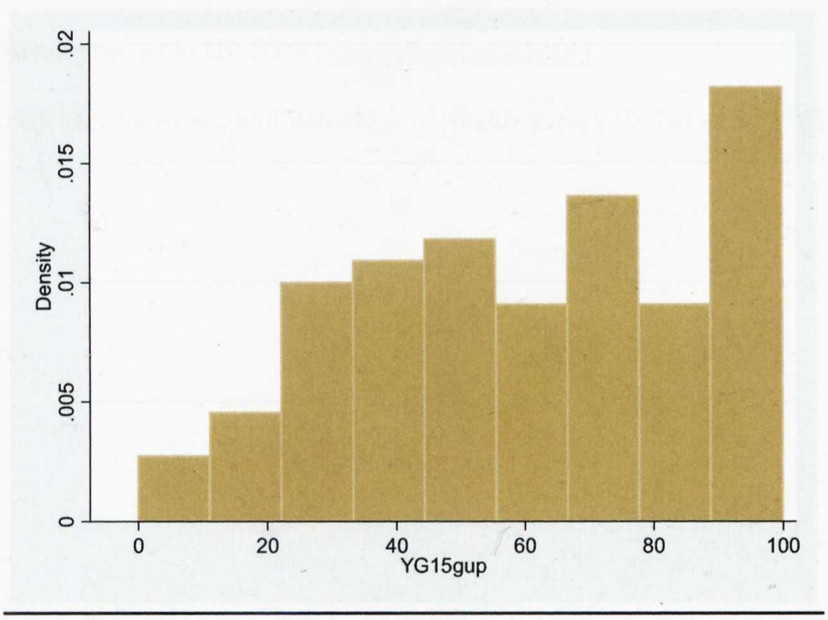
80% Income Gain (Years Traded)



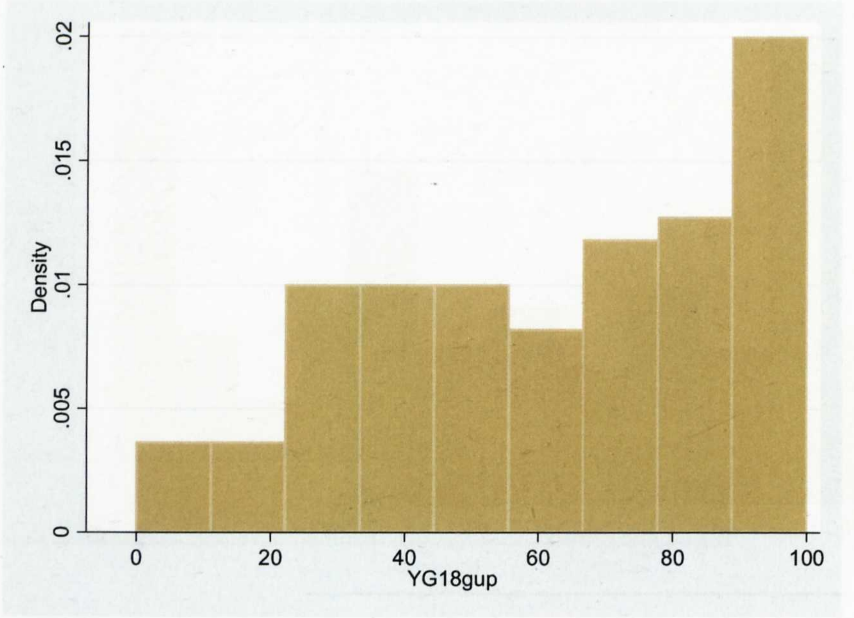
2 Years Gain (Percentage of Income Traded)



5 Years Gain (Percentage of Income Traded)



8 Years Gain (Percentage of Income Traded)



A 6.3 Notes on UK Statistics in Table 6.1

Gender: 2001 census (www.ons.gov.uk)

Age: As at mid-2008 (http://www.statistics.gov.uk/cci/nugget.asp?ID=6)

Educated beyond min school leaving age: Labour Force Survey. 72% is number obtaining GCSE's (or equivalent) or higher).

Degree: Labour Force Survey. Degree or higher Degree.

Employment: September –November 2009. (www.ons.gov.uk)

Net Monthly Income: Extrapolated from Annual Survey of Household Earnings (ASHE) 2009 (www.statistics.gov.uk). Median weekly pay for full time employees = £489. 4.33 * £489 = £2119. This value was inputted into www.thesalarycalculator.co.uk to give a monthly take home pay of £1622.46.

Children under 18 in the household: Census 2001 (www.ons.gov.uk)

Religion: British Social Attitudes Survey 2007 (www.britsocat.com). 10.3% were Christian (no denomination). These were categorised as Protestant.

Marital Status: As at mid-2008 (www.statistics.gov.uk). Co-habiting not included as a marital status.

Home Ownership: As at Q2 2009 (www.statistics.gov.uk).

EQ-VAS: UK Measurement and Valuation of Health Survey (Dolan *et al.* 1995; Kind *et al.* 1999)

Table A6.4 - Background Characteristics by Variant

		Variant A	Variant B	Variant C	Variant D	Chi^2 Test (p-values)
Number of Respondents		25	25	25	25	
Gender	Male	36%	44%	48%	40%	0.844
	Female	64%	56%	52%	60%	
Age	Average	42.44	36.16	44.84	37.72	
	SD	12.17	11.57	14.11	13.13	
	18-35	32%	56%	28%	40%	0.209
	36-50	44%	24%	36%	48%	
	51-69	24%	20%	36%	12%	
Educated beyond min school leaving age	Yes	96%	84%	80%	80%	0.338
	No	4%	16%	20%	20%	
Educated to Degree Level	Yes	72%	64%	64%	40%	0.112
	No	28%	36%	36%	60%	
Employment	Employed	60%	56%	40%	60%	0.434[1]
	Self-Employed	28%	24%	32%	12%	
	Retired	0%	0%	12%	4%	
	Housework	4%	12%	4%	4%	
	Student	0%	0%	0%	8%	
	Seeking Work	8%	12%	8%	8%	
	Other	0%	4%	4%	4%	
Net Own Monthly Income	<£430	8%	16%	20%	12%	0.742
	£430-£1300	44%	28%	24%	36%	
	£1300-£2200	32%	28%	40%	24%	
	>£2200	16%	28%	16%	28%	
Children	Yes	40%	40%	16%	40%	0.186
	No	60%	60%	84%	60%	
Religion	Protestant	4%	12%	8%	20%	0.693
	Roman Catholic	44%	52%	40%	36%	
	Atheist/Agnostic	40%	32%	48%	36%	
	Other	12%	4%	4%	8%	
Marital Status	Married	44%	20%	40%	36%	0.181
	Single	16%	48%	20%	40%	
	Co-habiting	20%	28%	32%	20%	
	Divorced	16%	0%	8%	4%	
	Widowed	4%	4%	0%	0%	
Home Ownership	Own/mortgage	66.67%	41.67%	48.00%	56.00%	0.337
	Rent	33.33%	58.33%	52.00%	44.00%	
Mean Self-Reported Health on the EQ-VAS		0.809	0.837	0.782	0.832	0.531

1. Employment treated as dichotomous, employed/self-employed vs the rest.

Table A6.5a - Background Characteristics by Traders and LG responders in the Income Loss questions				
		Total Traders[1]	LG responders	Chi^2 Test (p-values)
Number of Respondents		70	29	
Gender	Male	45.71%	31.03%	0.177
	Female	54.29%	68.97%	
Age	Average	38.51	43.93	
	SD	12.99	12.42	
	18-35	44.29%	27.59%	0.225
	36-50	37.14%	41.38%	
	51-69	18.57%	31.03%	
Educated beyond min school leaving age	Yes	85.71%	82.76%	0.709
	No	14.29%	17.24%	
Educated to Degree Level	Yes	60.00%	58.62%	0.899
	No	40.00%	41.38%	
Employment	Employed	61.43%	37.93%	**0.009**
	Self-Employed	24.29%	24.14%	
	Retired	1.43%	10.34%	
	Housework	4.29%	3.45%	
	Student	4.29%	6.90%	
	Seeking Work	1.43%	13.79%	
	Other	2.86%	3.45%	
Net Own Monthly Income	<£430	11.43%	17.24%	0.269
	£430-£1300	38.57%	20.69%	
	£1300-£2200	31.43%	31.03%	
	>£2200	18.57%	31.03%	
Children	Yes	34.29%	31.03%	0.755
	No	65.71%	68.97%	
Religion	Roman Catholic	11.43%	10.34%	0.169
	Protestant	50.00%	27.59%	
	Atheist/Agnostic	32.86%	55.17%	
	Other	5.71%	6.90%	
Marital Status	Married	35.71%	31.03%	**0.066**
	Single	30.00%	34.48%	
	Co-habiting	30.00%	13.79%	
	Divorced	2.86%	17.24%	
	Widowed	1.43%	3.45%	
Home Ownership[3]	Own/mortgage	49.28%	35.71%	0.179
	Rent	50.72%	64.29%	
Mean Self-Reported Health on the EQ-VAS[4]		0.817	0.806	0.779[5]

1. People who trade in all 3 of the EL questions. 2. Employment treated as dichotomous, employed/self-employed vs the rest. 3. Relevant sample size=97. One respondent lived with their parents and one respondent lived in co-op housing (both of these fell into the "rest of the sample"). 4. Relevant sample size=98. One respondent was unable to position Dead on the scale (trading and non-trading). 5. This was a t-test.

Table A6.5b - Background Characteristics by Traders and LG responders in the **Income Gain** questions				
		Total Traders[1]	Rest of the sample	Chi^2 Test (p-values)
Number of Respondents		55	44	
Gender	Male	45.45%	36.36%	0.362
	Female	54.55%	63.64%	
Age	Average	38.91	41.59	
	SD	12.36	13.75	
	18-35	41.82%	36.36%	0.555
	36-50	40.00%	36.36%	
	51-69	18.18%	27.27%	
Educated beyond min school leaving age	Yes	85.45%	84.09%	0.851
	No	14.55%	15.91%	
Educated to Degree Level	Yes	60.00%	59.09%	0.927
	No	40.00%	40.91%	
Employment	Employed	67.27%	38.64%	**0.001**[2]
	Self-Employed	23.64%	25.00%	
	Retired	0.00%	9.09%	
	Housework	5.45%	2.27%	
	Student	0.00%	11.36%	
	Seeking Work	1.82%	9.09%	
	Other	1.82%	4.55%	
Net Own Monthly Income	<£430	7.27%	20.45%	0.143
	£430-£1300	34.55%	31.82%	
	£1300-£2200	38.18%	22.73%	
	>£2200	20.00%	25.00%	
Children	Yes	29.09%	38.64%	0.317
	No	70.91%	61.36%	
Religion	Roman Catholic	10.91%	11.36%	0.843
	Protestant	40.00%	47.73%	
	Atheist/Agnostic	7.27%	4.55%	
	Other	41.82%	36.36%	
Marital Status	Married	32.73%	36.36%	0.129
	Single	27.27%	36.36%	
	Co-habiting	32.73%	15.91%	
	Divorced	3.64%	11.36%	
	Widowed	3.64%	0.00%	
Home Ownership[3]	Own/mortgage	50.00%	58.14%	0.425
	Rent	50.00%	41.86%	
Mean Self-Reported Health on the EQ-VAS[4]		0.790	0.845	0.134[5]

1. People who trade in all 3 of the IG questions. 2. Employment treated as dichotomous, employed/self-employed vs the rest. 3. Relevant sample size=97. One respondent lived with their parents and one respondent lived in co-op housing (both of these fell into the trading and non-trading category). 4. Relevant sample size=98. One respondent was unable to position Dead on the scale (trading and non-trading). 5. This was a t-test.

Table A6.5c - Background Characteristics by Traders and Non-traders in the Years Loss questions

		Traders	Total LG responders[1]	Chi^2 Test (p-values)
Number of Respondents		55	44	
Gender	Male Female	45% 55%	36% 64%	0.362
Age	Average SD 18-35 36-50 51-69	37.05 13.03 55% 27% 18%	43.91 12.04 21% 52% 27%	**0.002**
Educated beyond min school leaving age	Yes No	84% 16%	86% 14%	0.707
Educated to Degree Level	Yes No	51% 49%	70% 30%	**0.049**
Employment	Employed Self-Employed Retired Housework Student Seeking Work Other	60% 22% 0% 5% 7% 4% 2%	48% 27% 9% 2% 2% 7% 5%	0.410[2]
Net Own Monthly Income	<£430 £430-£1300 £1300-£2200 >£2200	13% 36% 36% 15%	14% 30% 25% 32%	0.200
Children	Yes No	73% 27%	41% 59%	0.153
Religion	Roman Catholic Protestant Atheist/Agnostic Other	13% 42% 5% 40%	9% 45% 39% 7%	0.929
Marital Status	Married Single Co-habiting Divorced Widowed	27% 33% 35% 4% 2%	43% 30% 14% 11% 2%	**0.084**
Home Ownership[3]	Own/mortgage Rent	38% 62%	70% 30%	**0.001**
Mean Self-Reported Health on the EQ-VAS[4]		0.811	0.818	0.838[5]

1. People who gave LG responses in all 3 of the YL questions. 2. Employment treated as dichotomous, employed/self-employed vs the rest. 3. Relevant sample size=97. One respondent lived with their parents and one respondent lived in co-op housing (both of these fell into the "rest of the sample"). 4. Relevant sample size=98. One respondent was unable to position Dead on the scale 5. This was a t-test.

Table A6.5d - Background Characteristics by Traders and Non-traders in the Years Gain questions				
		Total Traders[1]	LG responders	Chi^2 Test (p-values)
Number of Respondents		79	20	
Gender	Male	43%	35%	0.514
	Female	57%	65%	
Age	Average	39.72	41.60	
	SD	13.28	12.02	
	18-35	42%	30%	0.627
	36-50	37%	45%	
	51-69	21%	25%	
Educated beyond min school leaving age	Yes	86%	80%	0.498
	No	14%	20%	
Educated to Degree Level	Yes	59%	60%	0.967
	No	41%	40%	
Employment	Employed	49%	75%	0.170
	Self-Employed	27%	15%	
	Retired	5%	0%	
	Housework	5%	0%	
	Student	6%	0%	
	Seeking Work	5%	5%	
	Other	3%	5%	
Net Own Monthly Income	<£430	14%	10%	0.682
	£430-£1300	36%	25%	
	£1300-£2200	30%	35%	
	>£2200	20%	30%	
Children	Yes	32%	40%	0.479
	No	68%	60%	
Religion	Roman Catholic	10%	15%	**0.089**
	Protestant	48%	25%	
	Atheist/Agnostic	34%	60%	
	Other	8%	0%	
Marital Status	Married	31%	45%	0.295
	Single	33%	25%	
	Co-habiting	28%	15%	
	Divorced	5%	15%	
	Widowed	3%	0%	
Home Ownership[3]	Own/mortgage	53%	50%	0.800
	Rent	47%	50%	
Mean Self-Reported Health on the EQ-VAS[4]		0.832	0.744	**0.094**[5]

1. People who trade in all 3 of the IL questions. 2. Employment treated as dichotomous, employed/self-employed vs the rest. 3. Relevant sample size=97. One respondent lived with their parents and one respondent lived in co-op housing (both of these fell into the "rest of the sample"). 4. Relevant sample size=98. One respondent was unable to position Dead on the scale (trading and non-trading). 5. This was a t-test.

Table A6.6a - The effect of life satisfaction[1] on LG responding in the income loss questions				
	Analysed Sample	Total Traders	Rest of the sample	t-test (p-values)
Number of Respondents	99	70	29	
Mean Satisfaction with Health	5.74	5.76	5.69	0.778
Mean Satisfaction with Job[2]	4.67	4.69	4.61	0.813
Mean Satisfaction with Social Life	5.17	5.23	5.00	0.459
Mean Satisfaction with Leisure Time	4.89	4.87	4.93	0.855
Mean Satisfaction with Life Overall	5.39	5.36	5.47	0.667
Mean Satisfaction with Income	4.41	4.30	4.69	0.277

1 Life Satisfaction is on a scale from 1-7, where 1 represents 'not satisfied at all' and 7 represents 'completely satisfied'. 2. Relevant sample size=85.
Total Trader: traded in all three IL questions

Table A6.6b - The effect of life satisfaction[1] on LG responding in the income gain questions				
	Analysed Sample	Total Traders	Rest of the sample	t-test (p-values)
Number of Respondents	99	55	44	
Mean Satisfaction with Health	5.74	5.69	5.80	0.633
Mean Satisfaction with Job[2]	4.67	4.54	4.88	0.297
Mean Satisfaction with Social Life	5.17	5.20	5.11	0.761
Mean Satisfaction with Leisure Time	4.89	4.93	4.84	0.772
Mean Satisfaction with Life Overall	5.39	5.42	5.35	0.775
Mean Satisfaction with Income	4.41	4.33	4.52	0.553

1 Life Satisfaction is on a scale from 1-7, where 1 represents 'not satisfied at all' and 7 represents 'completely satisfied'. 2. Relevant sample size=85.
Total Trader: traded in all three IG questions

Table A6.6c - The effect of life satisfaction[1] on LG responding in the Years Loss questions				
	Analysed Sample	Total LG responders	Rest of the sample	t-test (p-values)
Number of Respondents	99	44	55	
Mean Satisfaction with Health	5.74	5.7	5.76	0.785
Mean Satisfaction with Job[2]	4.67	4.57	4.75	0.572
Mean Satisfaction with Social Life	5.17	5.09	5.22	0.654
Mean Satisfaction with Leisure Time	4.89	4.93	4.85	0.797
Mean Satisfaction with Life Overall	5.39	5.48	5.32	0.483
Mean Satisfaction with Income	4.41	4.86	4.05	**0.013**

1 Life Satisfaction is on a scale from 1-7, where 1 represents 'not satisfied at all' and 7 represents 'completely satisfied'. 2. Relevant sample size=85.
Total LG responder: gave an LG response in all three YL questions

Table A6.6d - The effect of life satisfaction[1] on LG responding in the Years Gain questions				
	Analysed Sample	Total Traders	LG responders	t-test (p-values)
Number of Respondents	99	79	20	
Mean Satisfaction with Health	5.74	5.81	5.45	0.182
Mean Satisfaction with Job[2]	4.67	4.74	4.42	0.401
Mean Satisfaction with Social Life	5.17	5.23	4.90	0.349
Mean Satisfaction with Leisure Time	4.89	5.00	4.45	0.134
Mean Satisfaction with Life Overall	5.39	5.41	5.30	0.696
Mean Satisfaction with Income	4.41	4.32	4.80	0.234

1 Life Satisfaction is on a scale from 1-7, where 1 represents 'not satisfied at all' and 7 represents 'completely satisfied'. 2. Relevant sample size=85.
Total Trader: traded in all three YG questions

Table A6.7a - The effect of income expectations on LG responding in the Income Loss Questions

		Full Sample	Total Traders	Rest of the Sample	t-test/chi² (p-values)
Number of Respondents		99	70	29	
Mean level below which income would be considered very bad (£ per month)		1457	1568	1191	0.176[1]
Mean Lower Bound of Income Sufficiency		2250	2375	1949	0.302[1]
Mean Upper Bound of Income Sufficiency		3266	3374	3004	0.496[1]
Mean level above which income would be considered excellent		5303	5660	4444	0.439[1]
Actual Income exceeds level below which income would be very bad	Yes	79.80%	81.43%	75.86%	
	No	20.20%	18.57%	24.14%	0.530[2]
Actual Income exceeds lower bound of income sufficiency	Yes	65.66%	67.14%	62.07%	
	No	34.34%	32.86%	37.93%	0.628[2]
Actual Income exceeds Upper Bound of Income Sufficiency	Yes	41.41%	42.86%	37.93%	
	No	58.59%	57.14%	62.07%	0.651[2]
Actual Income exceeds level above which Income would be considered excellent	Yes	19.19%	21.43%	13.79%	
	No	80.81%	78.57%	86.21%	0.380[2]

1. t-test. 2. Chi2
Total Trader: traded in all three IL questions

Table A6.7b - The effect of income expectations on LG responding in the Income Gain Questions					
		Full Sample	Total Traders	Rest of the Sample	t-test/chi² (p-values)
Number of Respondents		99	55	44	
Mean level below which income would be considered very bad (£ per month)		1457	1543	1351	0.583[1]
Mean Lower Bound of Income Sufficiency		2250	2208	2303	0.867[1]
Mean Upper Bound of Income Sufficiency		3266	2993	3606	0.419[1]
Mean level above which income would be considered excellent		5303	5627	4899	0.718[1]
Actual Income exceeds level below which income would be very bad	Yes	79.80%	85.45%	72.73%	
	No	20.20%	14.55%	27.27%	0.117[2]
Actual Income exceeds lower bound of income sufficiency	Yes	65.66%	70.91%	59.09%	
	No	34.34%	29.09%	40.91%	0.219[2]
Actual Income exceeds Upper Bound of Income Sufficiency	Yes	41.41%	45.45%	36.36%	
	No	58.59%	54.55%	63.64%	0.362[2]
Actual Income exceeds level above which Income would be considered excellent	Yes	19.19%	25.45%	11.36%	
	No	80.81%	74.55%	88.64%	**0.077**[2]

1. t-test 2. Chi²
Total Trader: traded in all three IG questions

Table A6.7c - The effect of income expectations on the decision to trade in the Years Loss Questions					
		Full Sample	Total LG responders	Rest of the Sample	t-test/chi^2 (p-values)
Number of Respondents		99	44	55	
Mean level below which income would be considered very bad (£ per month)		1457	1244	1628	0.243[1]
Mean Lower Bound of Income Sufficiency		2250	2061	2402	0.493[1]
Mean Upper Bound of Income Sufficiency		3266	3382	3172	0.753[1]
Mean level above which income would be considered excellent		5303	4786	5718	0.634[1]
Actual Income exceeds level below which income would be very bad	Yes	80%	80%	80%	
	No	20%	20%	20%	0.955[2]
Actual Income exceeds lower bound of income sufficiency	Yes	66%	68%	64%	
	No	34%	32%	36%	0.636[2]
Actual Income exceeds Upper Bound of Income Sufficiency	Yes	41%	41%	42%	
	No	59%	59%	58%	0.927[2]
Actual Income exceeds level above which Income would be considered excellent	Yes	19%	16%	22%	
	No	81%	84%	78%	0.458[2]

1. t-test. 2. Chi^2
Total LG responder: gave an LG response in all three YL questions

Table A6.7d - The effect of income expectations on the decision to trade in the Years Gain Questions

		Full Sample	Total Traders	LG responders	t-test/chi² (p-values)
Number of Respondents		99	79	20	
Mean level below which income would be considered very bad (£ per month)		1457	1521	1206	0.227[1]
Mean Lower Bound of Income Sufficiency		2250	2334	1919	0.292[1]
Mean Upper Bound of Income Sufficiency		3266	3404	2719	0.182[1]
Mean level above which income would be considered excellent		5303	5597	4145	0.321[1]
Actual Income exceeds level below which income would be very bad	Yes	80%	78%	85%	
	No	20%	22%	15%	0.517[2]
Actual Income exceeds lower bound of income sufficiency	Yes	66%	62%	80%	
	No	34%	38%	20%	0.130[2]
Actual Income exceeds Upper Bound of Income Sufficiency	Yes	41%	38%	55%	
	No	59%	62%	45%	0.167[2]
Actual Income exceeds level above which Income would be considered excellent	Yes	19%	84%	70%	
	No	81%	16%	30%	0.169[2]

1. t-test. 2. Chi²
Total Trader: traded in all three YG questions

Table A6.8a - Response to the feedback quesitons[1] by traders and LG responders in the Income Loss questions				
	Full Sample	Total Traders[2]	LG responders	t-test (p-values)
Number of Respondents	99	70	29	
The Visual Aid was helpful[3]	1.49	1.50	1.46	0.865
I could rely on my partners income[4]	3.21	3.16	3.34	0.622
If I didn't have children I would have answered differently[5]	2.85	2.93	2.65	0.551
Had in mind a minimum amount I could survive on	2.37	2.01	3.24	**0.001**
I consider it unethical to give up years of life for an increase in income	2.61	2.71	2.38	0.308
Want to be able to spend time with family and friends	1.61	1.71	1.34	**0.039**
I attempted to calculate the total lifetime income I would have in each scenario	3.61	3.30	4.34	**0.001**
Difficult to imagine my income falling by as much as specified in the questions	3.30	3.20	3.55	0.215

1. The feedback questions were coded 1-5, where 1 is strongly agree, and 5 is strongly disagree. 2. People who traded in all three of the EL questions. 3. Relevant sample size=98. One person did not answer this question. 3. 26 respondents had answered 'NA'. These were entered as 5 -Strongly disagree. 4. Sample size=59

Table A6.8b - Response to the feedback quesitons[1] by traders and LG responders in the Income Gain Questions

	Full Sample	Total Traders[2]	LG responders	t-test (p-values)
Number of Respondents	99	55	44	
The Visual Aid was helpful[3]	1.49	1.51	1.47	0.798
I could rely on my partners income[4]	3.21	3.00	3.48	0.157
If I didn't have children I would have answered differently[5]	2.85	2.85	2.84	0.976
Had in mind a minimum amount I could survive on	2.37	2.07	2.75	**0.023**
I consider it unethical to give up years of life for an increase in income	2.61	2.71	2.50	0.451
Want to be able to spend time with family and friends	1.61	1.76	1.41	**0.051**
I attempted to calculate the total lifetime income I would have in each scenario	3.61	3.24	4.07	**0.005**
Difficult to imagine my income falling by as much as specified in the questions	3.30	3.13	3.52	0.138

1. The feedback questions were coded 1-5, where 1 is strongly agree, and 5 is strongly disagree. 2. People who trade in all three CG questions. 3. Relevant sample size=98. One person did not answer this question. 3. 26 respondents had answered 'NA'. These were entered as 5 -Strongly disagree. 4. Sample size=59

Table A6.8c - Response to the feedback quesitons[1] by traders and LG responders in the Years Loss questions

	Full Sample	Traders	Total LG responders[2]	t-test (p-values)
Number of Respondents	99	55	44	
The Visual Aid was helpful[3]	1.49	1.49	1.49	0.988
I could rely on my partners income[4]	3.21	3.15	3.30	0.671
If I didn't have children I would have answered differently[5]	2.85	3.10	2.60	0.219
Had in mind a minimum amount I could survive on	2.37	2.16	2.64	0.117
I consider it unethical to give up years of life for an increase in income	2.61	2.65	2.57	0.758
Want to be able to spend time with family and friends	1.61	1.76	1.41	**0.051**
I attempted to calculate the total lifetime income I would have in each scenario	3.61	3.38	3.89	**0.095**
Difficult to imagine my income falling by as much as specified in the questions	3.30	3.07	3.59	**0.054**

1. The feedback questions were coded 1-5, where 1 is strongly agree, and 5 is strongly disagree. 2. Respondents who give LG responses in all three YL questions. 3. Respondents who Relevant sample size=98. One person did not answer this question. 3. 26 respondents had answered 'NA'. These were entered as 5 -Strongly disagree. 4. Sample size=59

Table A6.8d - Response to the feedback questions[1] by traders and LG responders in the Years Gain questions				
	Full Sample	Total Traders[2]	LG responders	t-test (p-values)
Number of Respondents	99	79	20	
The Visual Aid was helpful[3]	1.49	1.50	1.45	0.854
I could rely on my partners income[4]	3.21	3.32	2.80	0.240
If I didn't have children I would have answered differently[5]	2.85	2.79	3.00	0.666
Had in mind a minimum amount I could survive on	2.37	2.24	2.90	0.127
I consider it unethical to give up years of life for an increase in income	2.61	2.81	1.85	**0.001**
Want to be able to spend time with family and friends	1.61	1.66	1.40	0.298
I attempted to calculate the total lifetime income I would have in each scenario	3.61	3.39	4.45	**0.001**
Difficult to imagine my income falling by as much as specified in the questions	3.30	3.29	3.35	0.879

1. The feedback questions were coded 1-5, where 1 is strongly agree, and 5 is strongly disagree. 2. Respondents who traded in all three of the YG questions. 3. Relevant sample size=98. One person did not answer this question. 3. 26 respondents had answered 'NA'. These were entered as 5 -Strongly disagree. 4. Sample size=59

Table A6.9a- Background Characteristics by Over-traders in the Income Loss Questions				
		Over Traders[1]	Rest of the sample	Chi^2 Test (p-values)
Number of Respondents		31	68	
Gender	Male	48.39%	38.24%	0.342
	Female	51.61%	61.76%	
Age	Average	39.94	40.18	
	SD	15.47	11.83	
	18-35	51.61%	33.82%	**0.086**
	36-50	22.58%	45.59%	
	51-69	25.81%	20.59%	
Educated beyond min school leaving age	Yes	80.65%	86.76%	0.431
	No	19.35%	13.24%	
Educated to Degree Level	Yes	58.06%	60.29%	0.834
	No	41.94%	39.71%	
Employment	Employed	58.06%	52.94%	0.199[2]
	Self-Employed	19.35%	26.47%	
	Retired	6.45%	2.94%	
	Housework	9.68%	1.47%	
	Student	6.45%	4.41%	
	Seeking Work	0.00%	7.35%	
	Other	0.00%	4.41%	
Net Own Monthly Income	<£430	16.13%	11.76%	**0.070**
	£430-£1300	35.48%	32.35%	
	£1300-£2200	41.94%	26.47%	
	>£2200	6.45%	29.41%	
Children	Yes	16.13%	41.18%	**0.014**
	No	83.87%	58.82%	
Religion	Roman Catholic	12.90%	10.29%	0.459
	Protestant	48.39%	41.18%	
	Atheist/Agnostic	29.03%	44.12%	
	Other	9.68%	4.41%	
Marital Status	Married	29.03%	36.76%	0.295
	Single	38.71%	27.94%	
	Co-habiting	29.03%	23.53%	
	Divorced	0.00%	10.29%	
	Widowed	3.23%	1.47%	
Home Ownership[3]	Own/mortgage	36.67%	61.19%	**0.025**
	Rent	63.33%	38.81%	
Mean Self-Reported Health on the EQ-VAS[4]		0.790	0.825	0.373[5]

1. People who over-trade in any of the IL questions. 2. Employment treated as dichotomous, employed or unemployed vs the rest. 3. Relevant sample size=97. One respondent lived with their parents and one respondent lived in co-op housing (both of these fell into the trading and non-trading category). 4. Relevant sample size=98. One respondent was unable to position Dead on the scale 5. This was a t-test.

Table A6.9b- Background Characteristics by Over-traders in the Income Gain Questions				
		Over Traders[1]	Rest of the sample	Chi^2 Test (p-values)
Number of Respondents		17	82	
Gender	Male	58.82%	37.80%	0.109
	Female	41.18%	62.20%	
Age	Average	39.06	40.32	
	SD	13.87	12.89	
	18-35	47.06%	37.80%	0.375
	36-50	23.53%	41.46%	
	51-69	29.41%	20.73%	
Educated beyond min school leaving age	Yes	76.47%	86.59%	0.290
	No	23.53%	13.41%	
Educated to Degree Level	Yes	41.18%	63.41%	**0.089**
	No	58.82%	36.59%	
Employment	Employed	64.71%	52.44%	0.295[2]
	Self-Employed	23.53%	24.39%	
	Retired	0.00%	4.88%	
	Housework	5.88%	3.66%	
	Student	0.00%	6.10%	
	Seeking Work	5.88%	4.88%	
	Other	0.00%	3.66%	
Net Own Monthly Income	<£430	11.76%	13.41%	0.283
	£430-£1300	52.94%	29.27%	
	£1300-£2200	23.53%	32.93%	
	>£2200	11.76%	24.39%	
Children	Yes	11.76%	37.80%	**0.038**
	No	88.24%	62.20%	
Religion	Roman Catholic	17.65%	9.76%	0.470
	Protestant	52.94%	41.46%	
	Atheist/Agnostic	23.53%	42.00%	
	Other	5.88%	6.10%	
Marital Status	Married	23.53%	36.59%	0.332
	Single	35.29%	30.49%	
	Co-habiting	35.29%	23.17%	
	Divorced	0.00%	8.54%	
	Widowed	5.88%	1.22%	
Home Ownership[3]	Own/mortgage	25.00%	59.26%	**0.012**
	Rent	75.00%	40.74%	
Mean Self-Reported Health on the EQ-VAS[4]		0.721	0.834	**0.018**[5]

1. People who over-trade in any of the IG questions. 2. Employment treated as dichotomous, employed or unemployed vs the rest. 3. Relevant sample size=97. One respondent lived with their parents and one respondent lived in co-op housing (both of these fell into the trading and non-trading category). 4. Relevant sample size=98. One respondent was unable to position Dead on the scale 5. This was a t-test.

Table A6.9c- Background Characteristics by Under-Traders in the Years Loss Questions				
		Under Traders[1]	Rest of the sample	Chi^2 Test (p-values)
Number of Respondents		14	85	
Gender	Male	64%	38%	**0.061**
	Female	36%	62%	
Age	Average	39.29	40.24	
	SD	13.88	12.93	
	18-35	50%	38%	0.619
	36-50	36%	39%	
	51-69	14%	23%	
Educated beyond min school leaving age	Yes	79%	86%	0.480
	No	21%	14%	
Educated to Degree Level	Yes	43%	62%	0.168
	No	57%	38%	
Employment	Employed	72%	52%	0.494[2]
	Self-Employed	14%	26%	
	Retired	0%	5%	
	Housework	14%	2%	
	Student	0%	6%	
	Seeking Work	0%	6%	
	Other	0%	3%	
Net Own Monthly Income	<£430	7%	14%	**0.097**
	£430-£1300	43%	32%	
	£1300-£2200	50%	28%	
	>£2200	0%	26%	
Children	Yes	14%	36%	0.103
	No	86%	64%	
Religion	Roman Catholic	14%	11%	0.528
	Protestant	57%	41%	
	Atheist/Agnostic	22%	42%	
	Other	7%	6%	
Marital Status	Married	29%	36%	0.544
	Single	50%	28%	
	Co-habiting	14%	27%	
	Divorced	7%	7%	
	Widowed	0%	2%	
Home Ownership[3]	Own/mortgage	14%	59%	**0.002**
	Rent	86%	41%	
Mean Self-Reported Health on the EQ-VAS[4]		0.768	0.822	0.326[5]

1. People who under-trade in any of the YL questions. 2. Employment treated as dichotomous, employed or unemployed vs the rest. 3. Relevant sample size=97. One respondent lived with their parents and one respondent lived in co-op housing (both of these fell into the trading and non-trading category). 4. Relevant sample size=98. One respondent was unable to position Dead on the scale 5. This was a t-test.

Table A6.9d- Background Characteristics by Under-traders in the Years Gain Questions

		Under Traders[1]	Rest of the sample	Chi2 Test (p-values)
Number of Respondents		32	67	
Gender	Male	47%	61%	0.446
	Female	53%	39%	
Age	Average	37.69	41.25	
	SD	13.67	12.61	
	18-35	53%	33%	0.148
	36-50	28%	43%	
	51-69	19%	24%	
Educated beyond min school leaving age	Yes	84%	85%	0.928
	No	16%	15%	
Educated to Degree Level	Yes	62%	58%	0.684
	No	38%	42%	
Employment	Employed	53%	55%	0.911[2]
	Self-Employed	25%	24%	
	Retired	0%	6%	
	Housework	10%	2%	
	Student	6%	4%	
	Seeking Work	3%	6%	
	Other	3%	3%	
Net Own Monthly Income	<£430	9%	15%	0.163
	£430-£1300	47%	27%	
	£1300-£2200	31%	31%	
	>£2200	13%	27%	
Children	Yes	13%	43%	**0.002**
	No	87%	57%	
Religion	Roman Catholic	12%	11%	0.249
	Protestant	44%	43%	
	Atheist/Agnostic	32%	43%	
	Other	12%	3%	
Marital Status	Married	28%	37%	0.165
	Single	44%	25%	
	Co-habiting	25%	25%	
	Divorced	0%	11%	
	Widowed	3%	2%	
Home Ownership[3]	Own/mortgage	22%	67%	**0.000**
	Rent	78%	33%	
Mean Self-Reported Health on the EQ-VAS[4]		0.760	0.840	**0.056**[5]

1. People who over-trade in any of the YG questions. 2. Employment treated as dichotomous, employed or unemployed vs the rest. 3. Relevant sample size=97. One respondent lived with their parents and one respondent lived in co-op housing (both of these fell into the trading and non-trading category). 4. Relevant sample size=98. One respondent was unable to position Dead on the scale 5. This was a t-test.

Table A6.10a - The effect of life satisfaction[1] on the likelihood of overtrading in the Income Loss questions				
	Full Sample	Over Traders[2]	Rest of the sample	t-test (p-values)
Number of Respondents	99	31	68	
Mean Satisfaction with Health	5.74	5.65	5.78	0.606
Mean Satisfaction with Job[3]	4.67	4.69	4.66	0.924
Mean Satisfaction with Social Life	5.17	5.35	5.07	0.357
Mean Satisfaction with Leisure Time	4.89	5.00	4.84	0.599
Mean Satisfaction with Life Overall	5.39	5.29	5.43	0.587
Mean Satisfaction with Income	4.41	4.32	4.46	0.700

1 Life Satisfaction is on a scale from 1-7, where 1 represents 'not satisfied at all' and 7 represents 'completely satisfied'. 2. People who over-trade in any of the EI questions. 3. Relevant sample size=85.

Table A6.10b - The effect of life satisfaction[1] on the likelihood of overtrading in the Income Gain questions				
	Full Sample	Over Traders[2]	Rest of the sample	t-test (p-values)
Number of Respondents	99	17	82	
Mean Satisfaction with Health	5.74	5.35	5.82	**0.098**
Mean Satisfaction with Job[3]	4.67	4.60	4.69	0.850
Mean Satisfaction with Social Life	5.17	5.18	5.16	0.957
Mean Satisfaction with Leisure Time	4.89	4.94	4.88	0.822
Mean Satisfaction with Life Overall	5.39	5.29	5.41	0.691
Mean Satisfaction with Income	4.41	4.00	4.50	0.199

1 Life Satisfaction is on a scale from 1-7, where 1 represents 'not satisfied at all' and 7 represents 'completely satisfied'. 2. People who over-trade in any of the CG questions 3. Relevant sample size=85.

Table A6.10c - The effect of life satisfaction[1] on the likelihood of overtrading in the Years Loss questions

	Full Sample	Under Traders[2]	Rest of the sample	t-test (p-values)
Number of Respondents	99	14	85	
Mean Satisfaction with Health	5.74	5.14	5.84	0.107
Mean Satisfaction with Job[3]	4.67	4.75	4.66	0.832
Mean Satisfaction with Social Life	5.17	4.86	5.21	0.522
Mean Satisfaction with Leisure Time	4.89	4.29	4.99	0.108
Mean Satisfaction with Life Overall	5.39	4.64	5.51	**0.061**
Mean Satisfaction with Income	4.41	3.57	4.55	**0.008**

1 Life Satisfaction is on a scale from 1-7, where 1 represents 'not satisfied at all' and 7 represents 'completely satisfied'. 2. People who under-trade in any of the YL questions 3. Relevant sample size=85.

Table A6.10d - The effect of life satisfaction[1] on the likelihood of overtrading in the Years Gain questions

	Full Sample	Under Traders[2]	Rest of the sample	t-test (p-values)
Number of Respondents	99	32	67	
Mean Satisfaction with Health	5.74	5.53	5.84	0.229
Mean Satisfaction with Job[3]	4.67	4.54	4.73	0.589
Mean Satisfaction with Social Life	5.17	5.03	5.22	0.528
Mean Satisfaction with Leisure Time	4.89	4.81	4.93	0.706
Mean Satisfaction with Life Overall	5.39	5.13	5.51	0.141
Mean Satisfaction with Income	4.41	3.75	4.73	**0.005**

1 Life Satisfaction is on a scale from 1-7, where 1 represents 'not satisfied at all' and 7 represents 'completely satisfied'. 2. People who over-trade in any of the YG questions 3. Relevant sample size=85.

Table A6.11a - The effect of income expectations on the likelihood of overtrading in the **Income Loss** Questions					
		Full Sample	Over Traders	Rest of sample	t-test/chi^2 (p-values)
Number of Respondents		99	31	68	
Mean level below which income would be considered very bad (£ per month)		1457	1567	1407	0.744[1]
Mean Lower Bound of Income Sufficiency		2250	2105	2316	0.724[1]
Mean Upper Bound of Income Sufficiency		3266	2441	3642	**0.032**[1]
Mean level above which income would be considered excellent		5303	3038	6336	**0.039**[1]
Actual Income exceeds level below which income would be very bad	Yes	79.80%	77.42%	80.88%	
	No	20.20%	22.58%	19.12%	0.691[2]
Actual Income exceeds lower bound of income sufficiency	Yes	65.66%	61.29%	67.65%	
	No	34.34%	38.71%	32.35%	0.537[2]
Actual Income exceeds Upper Bound of Income Sufficiency	Yes	41.41%	38.71%	42.65%	
	No	58.59%	61.29%	57.35%	0.712[2]
Actual Income exceeds level above which Income would be considered excellent	Yes	19.19%	19.35%	19.12%	
	No	80.81%	80.65%	80.88%	0.978[2]
1. t-test. 2. Chi2					

Table A6.11b - The effect of income expectations on the likelihood of overtrading in the Income Gain Questions

		Full Sample	Over Traders	Rest of sample	t-test/chi[2] (p-values)
Number of Respondents		99	17	82	
Mean level below which income would be considered very bad (£ per month)		1457	2033	1338	0.420[1]
Mean Lower Bound of Income Sufficiency		2250	1638	2898	**0.049**[1]
Mean Upper Bound of Income Sufficiency		3266	2470	3663	0.114[1]
Mean level above which income would be considered excellent		5303	2842	5814	**0.031**[1]
Actual Income exceeds level below which income would be very bad	Yes	79.80%	88.24%	78.05%	
	No	20.20%	11.76%	21.95%	0.341[2]
Actual Income exceeds lower bound of income sufficiency	Yes	65.66%	52.94%	68.29%	
	No	34.34%	47.06%	31.71%	0.225[2]
Actual Income exceeds Upper Bound of Income Sufficiency	Yes	41.41%	29.41%	43.90%	
	No	58.59%	70.59%	56.10%	0.270[2]
Actual Income exceeds level above which Income would be considered excellent	Yes	19.19%	17.65%	19.51%	
	No	80.81%	82.35%	80.49%	0.859[2]

1. t-test. 2. Chi2

Table A6.11c - The effect of income expectations on the likelihood of under-trading in the Years Loss Questions

		Full Sample	Under Traders	Rest of the Sample	t-test/chi2 (p-values)
Number of Respondents		99	14	85	
Mean level below which income would be considered very bad (£ per month)		1457	1997	1369	0.547[1]
Mean Lower Bound of Income Sufficiency		2250	1358	2397	**0.002**[1]
Mean Upper Bound of Income Sufficiency		3266	1952	3482	**0.001**[1]
Mean level above which income would be considered excellent		5303	2240	5808	**0.006**[1]
Actual Income exceeds level below which income would be very bad	Yes	80%	79%	80%	
	No	20%	21%	20%	0.902[2]
Actual Income exceeds lower bound of income sufficiency	Yes	66%	64%	66%	
	No	34%	36%	34%	0.907[2]
Actual Income exceeds Upper Bound of Income Sufficiency	Yes	41%	36%	42%	
	No	59%	64%	58%	0.640[2]
Actual Income exceeds level above which Income would be considered excellent	Yes	19%	21%	19%	
	No	81%	79%	81%	0.819[2]

1. t-test. 2. Chi2

Table A6.11d - The effect of income expectations on the likelihood of over-trading in the Years Gain Questions

		Full Sample	Under Traders	Rest of the Sample	t-test/chi2 (p-values)
Number of Respondents		99	32	67	
Mean level below which income would be considered very bad (£ per month)		1457	1938	1228	0.183[1]
Mean Lower Bound of Income Sufficiency		2250	2738	2018	0.374[1]
Mean Upper Bound of Income Sufficiency		3266	3447	3179	0.777[1]
Mean level above which income would be considered excellent		5303	4537	5669	0.553[1]
Actual Income exceeds level below which income would be very bad	Yes	80%	72%	84%	
	No	20%	28%	16%	0.175[2]
Actual Income exceeds lower bound of income sufficiency	Yes	66%	53%	72%	
	No	34%	47%	28%	**0.070**[2]
Actual Income exceeds Upper Bound of Income Sufficiency	Yes	41%	22%	51%	
	No	59%	78%	49%	**0.006**[2]
Actual Income exceeds level above which Income would be considered excellent	Yes	19%	9%	24%	
	No	81%	91%	76%	**0.087**[2]

1. t-test. 2. Chi2

Table A6.12a - Response to the feedback quesitons[1] by over-traders in the Income Loss questions				
	Full Sample	Over Traders	Rest of Sample	t-test (p-values)
Number of Respondents	99	31	68	
The Visual Aid was helpful[2]	1.49	1.68	1.40	0.109
I could rely on my partners income[3]	3.21	3.16	3.24	0.839
If I didn't have children I would have answered differently[4]	2.85	3.12	2.74	0.453
Had in mind a minimum amount I could survive on	2.37	2.06	2.51	0.121
I consider it unethical to give up years of life for an increase in income	2.61	2.65	2.60	0.887
Want to be able to spend time with family and friends	1.61	2.03	1.41	**0.003**
I attempted to calculate the total lifetime income I would have in each scenario	3.61	3.19	3.79	**0.067**
Difficult to imagine my income falling by as much as specified in the questions	3.30	3.06	3.41	0.201

1. The feedback questions were coded 1-5, where 1 is strongly agree, and 5 is strongly disagree. 2. Relevant sample size=98. One person did not answer this question. 3. 26 respondents had answered 'NA'. These were entered as 5 -Strongly disagree. 4. Sample size=59

Table A6.12b - Response to the feedback quesitons[1] by over-traders in the Income Gain questions				
	Full Sample	Over Traders	Rest of Sample	t-test (p-values)
Number of Respondents	99	17	82	
The Visual Aid was helpful[2]	1.49	1.88	1.41	**0.033**
I could rely on my partners income[3]	3.21	3.18	3.22	0.924
If I didn't have children I would have answered differently[4]	2.85	3.50	2.75	0.285
Had in mind a minimum amount I could survive on	2.37	2.24	2.40	0.650
I consider it unethical to give up years of life for an increase in income	2.61	2.82	2.57	0.513
Want to be able to spend time with family and friends	1.61	2.06	1.51	**0.022**
I attempted to calculate the total lifetime income I would have in each scenario	3.61	2.94	3.74	**0.063**
Difficult to imagine my income falling by as much as specified in the questions	3.30	2.88	3.39	0.149

1. The feedback questions were coded 1-5, where 1 is strongly agree, and 5 is strongly disagree. 2. Relevant sample size=98. One person did not answer this question. 3. 26 respondents had answered 'NA'. These were entered as 5 -Strongly disagree. 4. Sample size=59

Table A6.12c - Response to the feedback quesitons[1] by over-traders in the Years Loss questions				
	Full Sample	Under Traders	Rest of the Sample	t-test (p-values)
Number of Respondents	99	14	85	
The Visual Aid was helpful[2]	1.49	1.86	1.43	**0.072**
I could rely on my partners income[3]	3.21	3.71	3.13	0.222
If I didn't have children I would have answered differently[4]	2.85	2.88	2.84	0.958
Had in mind a minimum amount I could survive on	2.37	2.21	2.40	0.667
I consider it unethical to give up years of life for an increase in income	2.61	2.57	2.62	0.904
Want to be able to spend time with family and friends	1.61	2.07	1.53	**0.091**
I attempted to calculate the total lifetime income I would have in each scenario	3.61	3.29	3.66	0.423
Difficult to imagine my income falling by as much as specified in the questions	3.30	3.21	3.32	0.760

1. The feedback questions were coded 1-5, where 1 is strongly agree, and 5 is strongly disagree. 2. Relevant sample size=98. One person did not answer this question. 3. 26 respondents had answered 'NA'. These were entered as 5 -Strongly disagree. 4. Sample size=59

Table A6.12d - Response to the feedback quesitons[1] by over-traders in the Years Gain questions				
	Full Sample	Under Traders	Rest of the Sample	t-test (p-values)
Number of Respondents	99	32	67	
The Visual Aid was helpful[2]	1.49	1.63	1.42	0.250
I could rely on my partners income[3]	3.21	3.59	3.03	0.110
If I didn't have children I would have answered differently[4]	2.85	2.81	2.86	0.917
Had in mind a minimum amount I could survive on	2.37	1.94	2.58	0.014
I consider it unethical to give up years of life for an increase in income	2.61	2.88	2.49	0.187
Want to be able to spend time with family and friends	1.61	1.91	1.46	**0.033**
I attempted to calculate the total lifetime income I would have in each scenario	3.61	3.34	3.73	0.235
Difficult to imagine my income falling by as much as specified in the questions	3.30	3.16	3.37	0.418

1. The feedback questions were coded 1-5, where 1 is strongly agree, and 5 is strongly disagree. 2. Relevant sample size=98. One person did not answer this question. 3. 26 respondents had answered 'NA'. These were entered as 5 -Strongly disagree. 4. Sample size=59

Table A6.13 -Number of years traded in the Income Loss and Income Gain questions excluding LG responders						
	Income Loss			**Income Gain**		
	20%	**50%**	**80%**	**20%**	**50%**	**80%**
N	70	87	89	55	74	80
Mean	1.55	2.76	4.84	1.24	1.52	1.88
SD	1.54	2.39	2.97	1.57	1.26	1.50
10th Percentile	0.16	0.50	1.00	0.17	0.25	0.22
Median	1.25	2.25	5.00	0.75	1.25	1.50
90th Percentile	3.00	6.75	9.75	2.50	3.00	3.63

Table A6.14 - MVQ values calculated at the individual level using the Income Loss and Income Gain questions (*excluding* LG responders *and* over-traders) based on both individual income and household income

		Income Loss			Income Gain		
		20%	**50%**	**80%**	**20%**	**50%**	**80%**
		(n=51)	(n=71)	(n=74)	(n=41)	(n=68)	(n=73)
Mean number of years traded		0.88	1.84	4.02	0.59	1.26	1.60
Mean Annual Individual Income (£)		20,826	19,237	19,511	21,804	21,004	19,664
Mean Annual Household Income (£)		36,427	33,464	32,200	39,197	35,386	33,974
Mean VAS Own Health		0.833	0.833	0.823	0.812	0.831	0.838
Value of 1 Quality Adjusted Life Year (Based on Individual Income, £)	Mean	**125,475**	**97,096**	**83,723**	**223,546**	**293,693**	**290,027**
	SD	237,801	132,457	179,596	492,561	659,917	620,826
	10th percentile	3,520	3,592	3,422	16,800	8,610	10,736
	Median	33,600	32,842	14,790	62,553	70,628	64,377
	90th Percentile	281,600	302,400	238,255	439,824	702,551	585,900
Value of 1 Quality Adjusted Life Year (Based on Household Income, £)	Mean	**228,394**	**156,861**	**128,111**	**305,246**	**493,584**	**566,178**
	SD	432,872	205,011	238,344	537,574	982,892	1,240,274
	10th percentile	3,826	7,040	3,911	19,800	9,567	15,842
	Median	64,330	65,436	30,814	92,062	88,713	130,042
	90th Percentile	787,500	472,500	437,400	765,000	2,081,250	2,247,243

Table A6.15 - MVQ values calculated at the individual level using the Income Loss and Income Gain questions (*excluding* LG responders) based on both individual income and household income (with a balanced panel)

		Equivalent Income Loss			Compensating Income Gain		
		20%	**50%**	**80%**	**20%**	**50%**	**80%**
		(n=50)	(n=50)	(n=50)	(n=50)	(n=50)	(n=50)
Mean number of years traded		1.75	3.27	5.33	1.32	1.83	2.37
Mean Annual Individual Income (£)		21,180	21,180	21,180	21,180	21,180	21,180
Mean Annual Household Income (£)		36,504	36,504	36,504	36,504	36,504	36,504
Mean VAS Own Health		0.791	0.791	0.791	0.791	0.791	0.791
Number of negative responses (truncated at zero)		18	15	13	14	6	6
Value of 1 Quality Adjusted Life Year (Based on Individual Income, £)	Mean	**97,203**	**61,027**	**36,224**	**148,690**	**238,803**	**222,015**
	SD	232,323	113,133	62,763	428,393	621,557	620,011
	10th Percentile	0	0	0	0	0	0
	Median	8,844	17,050	8,125	24,185	25,447	34,173
	90th Percentile	245,435	159,909	115,655	348,224	630,795	387,392
Value of 1 Quality Adjusted Life Year (Based on Household Income, £)	Mean	**145,393**	**106,888**	**68,353**	**209,667**	**402,711**	**370,223**
	SD	345,867	192,931	145,099	480,922	958,382	834,716
	10th Percentile	0	0	0	0	0	0
	Median	21,487	24,122	15,857	40,482	35,922	56,548
	90th Percentile	249,113	325,400	175,787	578,845	1,208,301	716,720

Table A6.16a - Mean number of years traded across the three Income Loss levels, by response to the life satisfaction questions[1]				
	Full Sample	Yes	No	t-test (p-values)
Satisfied with Health[2]	2.62 (99)	2.62 (68)	2.63 (31)	0.984
Satisfied with Job[3]	2.61 (85)	2.58 (48)	2.64 (37)	0.895
Satisfied with Social Life [2]	2.62 (99)	2.70 (45)	2.56 (54)	0.734
Satisfied with Leisure Time[2]	2.62 (99)	2.12 (37)	2.93 (62)	0.038
Satisfied with Life Overall[2]	2.62 (99)	2.43 (54)	2.86 (45)	0.308
Satisfied with Income[3]	2.62 (99)	2.05 (51)	3.24 (48)	**0.004**

1. The life satisfaction questions were coded 1-7, where 7 is very satisfied, and 1 is very unsatisfied. 2. Yes=6-7. No=1-5. 3. Yes=5-7. No=1-4.

Table A6.16b - Mean number of years traded across the three Income Gain levels, by response to the life satisfaction questions[1]				
	Full Sample	Yes	No	t-test (p-values)
Satisfied with Health[2]	1.11 (99)	1.00 (68)	1.37 (31)	0.169
Satisfied with Job[3]	1.14 (85)	1.18 (48)	1.08 (37)	0.718
Satisfied with Social Life[2]	1.11 (99)	1.14 (45)	1.09 (54)	0.865
Satisfied with Leisure Time[2]	1.11 (99)	0.82 (37)	1.29 (62)	**0.035**
Satisfied with Life Overall[2]	1.11 (99)	1.04 (54)	1.20 (45)	0.504
Satisfied with Income[3]	1.11 (99)	0.89 (51)	1.36 (48)	**0.053**

1. The life satisfaction questions were coded 1-7, where 7 is very satisfied, and 1 is very unsatisfied. 2. Yes=6-7. No=1-5. 3. Yes=5-7. No=1-4.

Table A6.16c - Mean amount of income given up to achieve a gain in years across the three Years Gain levels, by response to the life satisfaction questions[1]

	Full Sample	Yes	No	t-test (p-values)
Satisfied with Health[2]	58.60 (99)	58.46 (68)	58.91 (31)	0.945
Satisfied with Job[3]	59.51 (85)	59.32 (48)	59.76 (37)	0.941
Satisfied with Social Life[2]	58.60 (99)	57.07 (45)	59.88 (54)	0.615
Satisfied with Leisure Time[2]	58.60 (99)	62.59 (37)	56.22 (62)	0.245
Satisfied with Life Overall[2]	58.60 (99)	61.56 (54)	55.06 (45)	0.251
Satisfied with Income[3]	58.60 (99)	65.49 (51)	51.28 (48)	**0.010**

1. The life satisfaction questions were coded 1-7, where 7 is very satisfied, and 1 is very unsatisfied. 2. Yes=6-7. No=1-5. 3. Yes=5-7. No=1-4.

Table A6.17a - Mean number of years traded across the three Income Loss levels, by response to the feedback questions[1]				
	Full Sample	Yes	No	t-test (p-values)
The Visual Aid was helpful[2]	2.62 (99)	2.20 (65)	3.43 (34)	**0.008**
I could rely on my partners income[34]	2.62 (99)	2.49 (37)	2.70 (62)	0.607
If I didn't have children I would have answered differently[3]	2.51 (59)	2.59 (27)	2.45 (32)	0.785
Had in mind a minimum amount I could survive on2	2.62 (99)	2.76 (37)	2.54 (62)	0.606
I consider it unethical to give up years of life for an increase in income[3]	2.62 (99)	2.28 (49)	2.97 (50)	**0.095**
Want to be able to spend time with family and friends[2]	2.62 (99)	1.94 (60)	3.68 (39)	**0.000**
I attempted to calculate the total lifetime income I would have in each scenario[3]	2.62 (99)	3.09 (28)	2.44 (71)	0.135
Difficult to imagine my income falling by as much as specified in the questions[3]	2.62 (99)	2.71 (28)	2.59 (71)	0.783

1. The feedback questions were coded 1-5, where 1 is strongly agree, and 5 is strongly disagree. 2. Yes=1. No=2-5. 3. Yes=1-2. No=3-5. 4. 26 respondents had answered 'NA'. These were entered as 5 -Strongly disagree.

Table A6.17b - Mean number of years traded across the three Income Gain levels, by response to the feedback questions1				
	Full Sample	Yes	No	t-test (p-values)
The Visual Aid was helpful[2]	1.11 (99)	0.98 (65)	1.36 (34)	0.144
I could rely on my partners income[34]	1.11 (99)	1.28 (37)	1.01 (62)	0.308
If I didn't have children I would have answered differently[3]	1.01 (59)	0.71 (27)	1.27 (37)	**0.070**
Had in mind a minimum amount I could survive on[2]	1.11 (99)	1.05 (37)	1.15 (62)	0.703
I consider it unethical to give up years of life for an increase in income[3]	1.11 (99)	0.95 (49)	1.28 (50)	0.167
Want to be able to spend time with family and friends[2]	1.11 (99)	1.80 (39)	0.67 (60)	**0.000**
I attempted to calculate the total lifetime income I would have in each scenario[3]	1.11 (99)	1.48 (28)	0.97 (71)	**0.061**
Difficult to imagine my income falling by as much as specified in the questions3	1.11 (99)	1.27 (28)	1.05 (71)	0.470

1. The feedback questions were coded 1-5, where 1 is strongly agree, and 5 is strongly disagree. 2. Yes=1. No=2-5. 3. Yes=1-2. No=3-5. 4. 11 respondents had answered 'NA'. These were entered as 5 -Strongly disagree.

Table A6.17c - Mean amount of income given up to achieve a gain in years across the three Years Gain levels, by response to the feedback questions[1]

	Full Sample	Yes	No	t-test (p-values)
The Visual Aid was helpful[2]	58.60 (99)	62.60 (65)	50.97 (34)	**0.041**
I could rely on my partners income[34]	58.60 (99)	61.57 (37)	56.83 (62)	0.401
If I didn't have children I would have answered differently[3]	62.70 (59)	59.41 (27)	65.48 (32)	0.390
Had in mind a minimum amount I could survive on[2]	58.60 (99)	56.43 (37)	59.90 (62)	0.543
I consider it unethical to give up years of life for an increase in income3	58.60 (99)	64.53 (49)	52.79 (50)	**0.034**
Want to be able to spend time with family and friends[2]	58.60 (99)	66.62 (60)	46.26 (39)	**0.000**
I attempted to calculate the total lifetime income I would have in each scenario[3]	58.60 (99)	50.66 (28)	61.73 (71)	**0.057**
Difficult to imagine my income falling by as much as specified in the questions[3]	58.60 (99)	60.84 (28)	57.72 (71)	0.587

1. The feedback questions were coded 1-5, where 1 is strongly agree, and 5 is strongly disagree. 2. Yes=1. No=2-5. 3. Yes=1-2. No=3-5. 4. 26 respondents had answered 'NA'. These were entered as 5 -Strongly disagree.

Table A6.18a - The effect of background characteristics, life satisfaction and feedback on the number of years traded in the Income Loss questions

	20% Income loss	50% income loss	80% income loss	Pooled (Long Dataset)
Gender (Male=1, Female=0)	0.751	0.013	-0.294	-0.086
Age	**-0.237***	**-0.244***	**-0.281****	**-0.207***
Age2	**0.003***	0.003	**0.003***	0.002
Educated beyond min school leaving age	0.445	1.380	0.545	0.751
Have a degree	-0.689	-0.360	0.373	-0.268
Married=1, Other=0	0.101	0.415	-0.538	0.205
Atheist/Agnostic=0, Religious=1	-1.012	**-1.033***	-0.750	**-1.057****
Employed=1, Not employed=0	**2.325****	0.547	0.010	0.622
Own Home/Mortage=1, Other=0	-0.375	0.131	-0.532	0.015
Income	-0.352	-0.099	0.111	-0.003
I can rely on my partner's income	0.517	0.727	**1.361***	0.556
Have children under18 in the house	-0.120	-0.919	-0.170	-0.692
Thought about family and friends	**-1.102****	**-1.533*****	**-1.606****	**-1.290*****
VAS Own Health	0.048	0.004	0.035	0.053
VAS Own Health Squared	-0.000	0.000	-0.000	-0.000
Had in mind a minimum amount I could survive on	-0.781	-0.413	0.721	0.051
Unethical to trade years for income	-0.041	-0.378	-0.373	-0.366
Aid was helpful	-0.766	**-0.980***	**-1.414****	**-0.884****
Attempted to calculate lifetime income	**1.056***	**0.831***	0.655	**0.873****
Difficult to imagine my income falling by as much as specified in the quesitons	0.172	0.552	0.200	0.332
I am satisfied with my health	-0.179	-0.112	0.911	0.198
I am satisfied with life overall	-0.514	0.110	0.050	-0.221
I am satisfied with my income	**-1.179****	**-1.542*****	-0.462	**-1.004****
Variant BADC	-0.482	-0.287	0.423	-0.276
Variant CDAB	0.012	-0.062	1.038	0.059
Variant DCBA	-0.022	-0.422	-0.522	-0.553
50% income loss vs 20% income loss				**1.504*****
80% income loss vs 20% income loss				**3.020*****
Number of Observations	98	98	98	294
Pseudo R^2	0.194	0.203	0.213	0.238
Link Test	0.069	0.474	0.304	0.014
Akaike Information Criterion	2.809	2.676	2.664	2.286

* Significant at 10% level ** Significant at 5% level *** Significant at 1% level
Dependent Variable categories: 20% Income Loss: 0 yrs traded=0; 0-0.5 yrs traded=1; 0.5-2 yrs traded=2; .2 yrs traded=3.
50% Income Loss: 0yrs traded=0; 0-2 yrs traded=1; 2-3 yrs traded=2; >3 yrs traded=3.
80% Income Loss: 0 yrs traded=0; 0-3 yrs traded=1; 3-6yrs traded=2; >6 yrs traded=3
Pooled: 0 yrs traded=0; 0-1 yrs traded=1; 1-3 yrs traded=2; >3 yrs traded=3.

Table 6.18b - The effect of background characteristics, life satisfaction and feedback on the number of years traded in the Income Gain questions				
	20% Income gain	50% income gain	80% income gain	Pooled (Long Dataset)
Gender (Male=1, Female=0)	0.368	-0.517	-0.376	0.174
Age	-0.196	**-0.379***	**-0.446*****	**-0.279***
Age2	0.002	**0.004***	**0.005****	**0.003***
Educated beyond min school leaving age	-0.005	1.175	0.526	0.370
Have a degree	-0.145	-0.644	**-0.946***	-0.630
Married=1, Other=0	-0.150	-0.176	-0.404	-0.343
Atheist/Agnostic=0, Religious=1	-0.621	-0.756	-0.915	-0.598
Employed=1, Not employed=0	**2.515****	0.886	**1.700****	**1.544****
Own Home/Mortage=1, Other=0	0.084	0.639	0.268	0.224
Income	0.207	0.430	0.275	0.275
I can rely on my partner's income	0.672	**1.115***	**1.546*****	**1.069****
Have children under18 in the house	-0.468	-0.742	-0.873	-0.444
Thought about family and friends	**-1.398****	**-1.726*****	**-1.663*****	**-1.594*****
VAS Own Health	-1.119	-0.066	-0.073	-0.065
VAS Own Health Squared	0.001	0.001	0.001	0.001
Had in mind a minimum amount I could survive on	-0.270	-0.561	**-0.962****	**-0.686***
Unethical to trade years for income	0.333	-0.544	-0.077	-0.077
Aid was helpful	-0.416	-0.429	-0.275	-0.338
Attempted to calculate lifetime income	**1.598****	**1.353*****	**1.488*****	**1.247*****
Difficult to imagine my income falling by as much as specified in the questions	-0.207	-0.306	-0.143	-0.164
I am satisfied with my health	-0.458	**-1.102***	**-1.126****	**-0.892***
I am satisfied with life overall	0.320	-0.039	-0.741	0.026
I am satisfied with my income	-0.891	**-1.167***	-0.772	**-0.970***
Variant BADC	-1.017	-0.624	-0.364	-0.792
Variant CDAB	-0.041	0.133	0.206	-0.209
Variant DCBA	0.478	0.242	-0.224	-0.115
50% income gain vs 20% income gain				**1.113*****
80% income loss vs 20% income gain				**1.727*****
Number of Observations	98	98	98	294
Pseudo R^2	0.219	0.187	0.229	0.033
Link Test	0.902	0.424	0.439	0.197
Akaike Information Criterion	2.184	2.833	2.716	2.417

Values Presented are coefficients

* Significant at 10% level ** Significant at 5% level *** Significant at 1% level

Dependent variable categories: 20% Income Gain: 0 yrs traded=0; 0-1 yrs traded=1; >1yr traded =2

50% Income Gain: 0 yrs traded=0; 0-0.5 yrs traded=1; 0.5-1.5 yrs traded=2; >1.5 yrs=3

80% Income Gain: 0 yrs traded=0; 0-1 yrs traded=1; 1-2 yrs traded =2; >2yrs=3

Pooled: 0 yrs traded=0; 0-0.5 yrs traded=1; 0.5-2 yrs traded=2; >2 yrs traded=3

Table A6.18c - The effect of background characteristics, life satisfaction and feedback on the amount of income given up in the Years Gain questions

	2 Equivalent Years Gain	5 Equivalent Years Gain	8 Equivalent Years Gain	Pooled (Long Dataset)
Gender (Male=1, Female=0)	0.669	0.056	-0.325	0.060
Age	0.246	**0.305***	0.244	0.242
Age2	-0.002	-0.003	-0.003	-0.003
Educated beyond min school leaving age	-0.755	-0.993	-0.829	-1.071
Have a degree	-0.879	-0.490	-0.099	-0.397
Married=1, Other=0	-0.095	0.249	-0.233	-0.113
Atheist/Agnostic=0, Religious=1	0.684	**1.211****	**1.009***	**0.906***
Employed/Self-Employed=1, Other=0	0.007	-0.031	0.463	0.211
Own Home/Mortage=1, Other=0	0.502	0.339	0.146	0.326
Income	-0.112	-0.182	-0.188	-0.145
I can rely on my partner's income	0.080	0.037	-0.634	-0.428
Have children under18 in the house	0.712	0.855	**1.456***	**0.989***
Thought about family and friends	**1.725*****	**1.393****	**0.903***	**1.220*****
VAS Own Health	0.025	0.001	0.020	0.042
VAS Own Health Squared	-0.000	-0.000	-0.000	-0.000
Had in mind a minimum amount I could survive on	-0.094	0.037	-0.090	0.033
Unethical to trade years for income	0.375	**0.866***	**1.257****	0.659
Aid was helpful	0.619	0.720	0.216	0.580
Attempted to calculate lifetime income	**-1.582*****	**-1.311****	**-1.267****	**-1.199*****
Difficult to imagine my income falling by as much as specified in the questions	**1.072***	0.559	0.561	0.682
I am satisfied with my health	-0.438	0.064	-0.150	-0.114
I am satisfied with life overall	0.364	0.007	0.288	0.297
I am satisfied with my income	0.180	**1.247****	**1.422*****	0.685
Variant BADC	0.494	1.166	0.535	0.661
Variant CDAB	-0.237	0.021	0.404	0.047
Variant DCBA	0.448	0.796	**1.160****	0.777
5 year gain vs 2 year gain				**0.816*****
8 year gain vs 2 year gain				**1.014*****
Number of Observations	98	98	98	294
Pseudo R^2	0.191	0.220	0.209	0.177
Link Test	0.716	0.405	0.652	0.158
Akaike Information Criterion	2.745	2.698	2.653	2.422

* Significant at 10% level ** Significant at 5% level *** Significant at 1% level Dependent variable categories: 2 Years Gain: 0% of income traded=0; 0-25% =1; 25%-50%=2; 50%-95%=3; >95%=4 5 Years Gain: 0% of income traded=0, 0-35%=1; 35%-70%=2; 70%-99.25%=3; >99.25%=4 8 years Gain:0% of income traded=0, 0-35%=1; 35%-80%=2; 80%-99.25%=3; >99.25%=4 Pooled: 0% of income traded=0, 0-30%=1; 30%-70%=2; 70%-99.25%=3; >99.25%=4