The University Of Sheffield.

A mixed methods investigation of methods of valuing health: are preferences over health states matters of taste, complete, and informed?

By:

Milad Karimi

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Date of submission: October 2015

The University of Sheffield
Faculty of Medicine, Dentistry and Health
School of Health and Related Research

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AQoL – The Assessment of Quality of Life

BTD – Better than dead

CEA – Cost Effectiveness Analysis

CLAMES - the Classification and Measurement System of Functional Health

CUA – Cost Utility Analysis

DCE – Discrete Choice Experiment

ICC – Intraclass correlation coefficient

ICER – Incremental Cost Effectiveness Ratio

MIC – Multi Instrument Comparison study

MVH – the Measurement and Valuation of Health

NICE - The National Institute for Health and Care Excellence

PTO – Person Trade Off

QALY – Quality Adjusted Life Year

QoL – Quality of Life

SD – Standard Deviation

SDQ – Sheffield Dignity Questionnaire

SG – Standard Gamble

TTO – Time Trade Off

VAS – Visual Analogue Scale

WTD – Worse than dead

Acknowledgements

Many people have provided me with support during the different stages of this PhD. I would like to thank my supervisors, John Brazier and Suzy Paisley. They have always been ready to help and their knowledge and experience has been invaluable. Both John and Suzy have been tremendously supportive and positive throughout the whole experience and I thank them for this.

I would like to thank staff members at ScHARR who have helped me over the last few years, Brendan Mulhern, Donna Rowen, Tracy Young, Emma Knowles, Louise Brereton, Hasan Basarir, Simon Palfreyman, and others. All have been very willing to answer any questions I had.

This research would not have been possible without the respondents, and the various volunteer and community organization in Sheffield who assisted in recruitment. I would also like to thank all the participants who took part in the pilot studies.

I must thank family and friends. Sharing an office with fellow PhD students (especially Sarah, Sabine, Sam, Ji Hee, and Laura) has made the journey much easier. My parents have always been incredibility supportive throughout this PhD.

The studentship I received from the University of Sheffield made it possible for me to enroll in the PhD and I thank them for funding this PhD study.

Conference papers based on this PhD

Based on the background chapters of this PhD:

The capability approach and health economics: are we applying it correctly?

Karimi M, Brazier J, Paisley S. Health Economists Study Group (HESG) Meeting. January 2013. University of Exeter.

The Capability Approach - What does it mean for health economics? Karimi M, Brazier J, Basarir H. 2013 EuroQol Group Scientific Plenary. September 2013. Montreal, Canada.

Based on study one of this PhD:

How do members of the general public value health states? Karimi M, Brazier J, Paisley S. 2015 EuroQol Group Scientific Plenary. September 2015. Krakow, Poland.

Abstract

Health economists use choice-based methods to value health states. These choice-based methods make assumptions about preferences. Three assumptions were investigated in this thesis, namely that preferences are: purely matters of taste, complete, and informed. Violations of these assumptions threaten the validity of choice-based methods. The aim of this thesis is to investigate the appropriateness of these three assumptions.

This PhD used a sequential mixed methods design with three studies. Qualitative interviews with a think-aloud protocol were undertaken to investigate whether preferences are purely matters of taste. A mixed methods study was conducted to test completeness by investigating the effect of reflection and deliberation on health state preferences. A quantitative study tested whether preferences are informed.

Three key findings were made. First, preferences over health are not purely matters of taste but depend on beliefs about how ill health affects an individual on domains such as enjoyment, independence, and relationships. Second, preferences were not shown to be incomplete because reflection and deliberation did not change aggregate mean health state values. Although individuals are uncertain about their preferences, reflection and deliberation does not seem to systematically alter their preferences. Third, preferences may not be informed because participants' beliefs about the consequences of ill health do not conform to the experience of patients in states of ill health.

This thesis contributes to knowledge about the role of beliefs in health state valuation, the effect of deliberation and reflection on health state preferences, and whether preferences are informed. Methodological contributions include developing a method of determining whether preferences are informed and the application of mixed methods.

A key implication of this PhD is that because preferences over health states are not entirely informed, choice-based methods of valuing health may provide sub-optimal policy recommendations. Recommendations for further research include implementing the methods in this thesis in a larger study and testing the effect of providing information about the effects of ill health to individuals valuing health.

Chapter 1

Introduction

Lynparza, an ovarian cancer drug, is not recommended for patients in England (NICE, 2015). The negative recommendation is not because Lynparza is ineffective. Lynparza is not recommended because it is too costly compared to its effectiveness (NICE, 2015). Lynparza is therefore not good value for money (NICE, 2015). In public health care systems, value for money calculations are made because governments have limited resources (Brazier et al., 2007, pp.7-8). Ideally, the resources employed in public health care systems should be employed effectively and efficiently (Brazier et al., 2007, pp.7-8). Different methods are available to governments to ensure that spending on health care treatment is effective and efficient. One method is the use of economic evaluation, where the costs and benefits of intervention are compared (Drummond et al., 2005, p.8). Only if the intervention provides enough benefits compared to the costs will the intervention be accepted (Drummond et al., 2005, p.8).

In England, the National Institute for Health and Care Excellence (NICE) is responsible for providing economic evaluations of some health technologies (Brazier et al., 2007, p.8; NICE, 2013a). When undertaking economic evaluations, NICE assesses the 'cost-effectiveness' of the technologies (NICE, 2013a). The practice of cost-effectiveness assessment is established but there remains debate about many aspects of cost-effectiveness methods (NICE, 2013a).

One of the debates is on methods used to value the benefits of health interventions. A health intervention can provide benefits on length of life, quality of life, or both (Brazier et al., 2007, p.26). The quality-adjusted life year (QALY) expresses both dimensions of benefit in one measure (Brazier et al., 2007, p.26). NICE uses the QALY

as a measure of benefit when assessing the cost-effectiveness of an intervention, hence the calculation of QALYs resulting from an intervention is an important aspect of performing cost-effectiveness (Drummond et al., 2005, p.173; NICE, 2013a). One QALY represents one year of full health (Drummond et al., 2005, p.175). To calculate the QALYs resulting from time in less than full health, one needs to know the 'quality weight' associated with those years (Drummond et al., 2005, p.174). Various methods can be used to determine quality weights for health states (Brazier et al., 2007, pp.83-94).

The type of methods recommended by NICE for determining quality weights are referred to as choice-based methods (NICE, 2013a). The validity of the choice-based methods depends on several assumptions (Torrance and Feeny, 1989; Hausman, 2006). This thesis examines three of those assumptions. In doing so, it hopes to contribute to the debate about the validity of choice-based methods for valuing health and about the policy recommendations these methods help provide.

1.1 Preferences and health state valuation

Choice-based methods of valuing health are so called because they rely on choices of individuals, often choices of members of the general public (Drummond et al., 2005, p.174; Brazier et al., 2007, pp.84-118). These choices are then meant to reveal the preferences of those individuals (Drummond et al., 2005, p.145). The quality weights are measured on a scale anchored at zero for being dead and one for being in full health, with values below zero used for states considered worse than being dead (Drummond et al., 2005, pp.174-175). Various choice-based methods exist to place heath states on a scale with the anchors of dead (0) and full health (1). Examples of choice-based methods of valuing health are the time trade off (TTO), the standard gamble (SG), or Discrete Choice Experiments (DCEs) (Torrance and Change, 1986;

Brazier et al., 2007, pp.87-93). In all these tasks participants are asked to make choices. For example, in the TTO participants are asked to choose between a certain number of years in a heath state compared to equal or fewer number of years in full health (Brazier et al., 2007, p.91). The participants' choices reveal their preferences and are used to calculate the value of health states. Preferences of individuals are thus central to the valuation task.

The use of preferences in health economics is largely derived from normative economics (Hausman, 2006; Morris et al., 2007, p.210). A dominant perspective in normative economics is welfarism or welfare economics (Morris et al., 2007, p.210). An important theory of welfarism is the preference satisfaction theory of welfare (Hausman, 2006; Brouwer et al., 2008; Beckerman, 2011, p.39). The preference satisfaction theory of welfare argues that 'the good' is the degree to which preferences of an individual are satisfied. Individuals are assumed to have preference orderings of states of the world and the achievement of higher orderings represents more welfare (Brouwer et al., 2008; Hausman and McPherson, 2009). The task for the health economist is to design an elicitation method to measure these preference orderings. Although non-welfarism is important in health economics, in practice the methods for valuing health still rely on preference (Brouwer et al., 2008). Investigating the assumptions of the preference satisfaction theory of welfare therefore is important for assessing the validity of choice-based methods of valuing health.

1.2 Assumptions about preferences

Three assumptions of the preference satisfaction theory of welfare are investigated in this thesis. The first two are explicitly stated, while the last is often implicitly held. The first assumption is that of complete preferences (Fischhoff, 1991; Morris et al., 2007, p.24; Beckerman, 2011, p.36). Individuals have complete preferences if they have

a preference ordering for all alternatives in the world. If given a choice between alternatives A and B, they will either prefer A to B, B to A, or be indifferent between the two (Hausman, 2012b, p.14). The completeness assumption suggests that individuals have existing preference orderings over a wide range of domains, including health states. This assumption may explain why typical health valuation tasks are one-off exercises, which respondents complete on their own without guidance on developing their preferences.

The second assumption is that the rationality of preferences is enough as a quality criterion in welfare economics. Economists agree that preferences should be rational, where rationality is determined by criteria such as consistency and transitivity (Sen, 2002, p.ch3; Morris et al., 2007, p.24; Beckerman, 2011, p.39). An example of consistency is that if A is preferred to B (from a choice set consisting of A and B), the addition of choice C to the set should not change the preference ordering between A and B (Sen, 1993). An individual's preferences are transitive if given that A is preferred to B and B is preferred to C, A is also preferred to C (Sen, 1993). Rationality, when framed as consistency and transitivity, is only a matter of "internal coherence" because only the relation of the preference ordering itself is needed to assess rationality, and no external information is required. (Kahneman, 1994; Sen, 2002, p.ch3).

The third assumption is that preferences are primarily matters of taste. Although this assumption is not a formal assumption of the preference satisfaction theory of welfare, sometimes this assumption is accepted implicitly (Hausman, 2011). Tastes are gut feelings and do not depend on any beliefs (Hausman, 2011). One cannot rationally argue about tastes (Stigler and Becker, 1977). When preferences are treated like tastes it means that rational discussion and investigation of preferences are not needed or perhaps even impossible (Hausman, 2006).

Challenges to the assumptions

These three assumptions have been challenged in the literature. First, the preference construction theory suggests that preferences over some domains may not be complete (Payne et al., 1992; Slovic, 1995). Researchers should not always assume that individuals have well-articulated values (Fischhoff, 1991). Second, rationality may not be a sufficient criterion of the quality of preferences. It is possible for consistent and transitive preferences to be 'wrong' because individual's can have external goals or objectives that their rational preferences can fail to satisfy (Sen, 2002, p.20). Preferences should not only be rational but individuals should have access to relevant information (Harsanyi, 1977). Preferences should be rational, but also "informed" (Gold et al., 1996, pp.99-100; Brazier et al., 2005). Third, some preferences are not matters of taste (Hausman, 2011). They involve complex cognitive processes and depend on factual beliefs about the world (Hausman, 2006). As opposed to matters of taste, one can discuss and criticise these preferences.

If these challenges to the assumptions are valid in the field of health state valuation, then the assumptions underlying choice-based methods may be violated in health state valuation. The extent to which these three assumptions are violated has implications for the validity of choice-based methods for valuing health. The aim of this thesis is to test the three assumptions and assess the implications of potential violations of the assumptions for methods of valuing health.

1.3 Rationale for investigating choice-based methods for valuing health

Investigating the assumptions is important because QALY values derived from choice-based tasks play an important role in determining resource allocation of health care

treatments. QALY values help determine resource allocation because they are used to assess the cost-effectiveness of health interventions. Recommendations made by NICE, and other agencies such as the Scottish Medicines Consortium, are partly based on the cost-effectiveness results (Cairns, 2006). The example of Lynparza not being recommended for patients in England shows the real consequences of this process (NICE, 2015).

In discussing the appropriateness of different assumptions Fischhoff (1991) suggests one should consider the severity of the consequences caused by operating on the wrong assumption. Fischhoff (1991) argues, "…one might not hire a survey researcher to study how acutely ill individuals evaluate alternative medical procedures, nor might one hire a philosopher to lead consumers through the intricacies of evaluating alternative dentifrices." Incorrect methods for valuing health may result in wrong decisions about which treatments to recommend for the NHS and its patients. The importance of these decisions makes research into the methods worthwhile.

1.4 Aims and objectives

The aim of this thesis is to investigate three assumptions of choice-based methods for valuing health. Three objectives are formulated in Chapter 2 to investigate each assumption:

1) To determine whether preferences can be assumed to be complete this PhD will investigate the role of reflection and deliberation (including Multiple Criteria Decision Analysis) in health state valuation
2) To understand whether preferences are informed this PhD will develop and implement methods to assess whether preferences over health states are informed

3) To understand whether preferences are purely matters of tastes this PhD will investigate the judgements, cognitive processes, and beliefs underlying preferences over health states

The three objectives are based on theoretical arguments presented in Chapter 2 and a literature review presented in Chapter 3. The link between the assumptions and the objectives are explained in detail in those two chapters. Achieving these objectives will improve existing knowledge on the validity of choice-based methods of valuing health.

1.5 Structure of thesis

This thesis consists of nine chapters. The next chapter, Chapter 2, provides the theoretical background and justification for this PhD. It argues that choice-based methods for valuing health are based on the preference satisfaction theory of welfare. It then describes in more detail the three assumptions of the theory and considers challenges to those assumptions and the appropriateness of the assumptions in the health care field. Based on the three assumptions, three research objectives are developed and described.

Chapter 3 reports a literature review relating to the three research objectives. The literature is summarised, discussed, and research gaps are highlighted. In Chapter 4 the study design and methods for reaching the research objectives are described.

Chapters 5 to 8 report on the empirical work done during this PhD. Chapter 5 describes a qualitative study investigating how members of the general public value health using two choice-based methods. The focus is on investigating the beliefs which individuals consider when valuing health.

Chapter 6 reports on study two, which investigates the effect of a structured reflection and deliberation exercise on individuals' health state preferences. The design of the exercise is described and justified, as well as the role of the findings from study

one (reported in Chapter 5) in designing the reflection and deliberation exercise. Health state values elicited before and after the exercise are compared quantitatively to determine the effect of reflection and deliberation on health state values.

The findings of study two in Chapter 6 had to be investigated further to draw accurate implications. A mixed methods study was conducted (reported in Chapter 7) for this reason. Audio recordings of the reflection and deliberation exercise were analysed qualitatively and compared with quantitative findings using a triangulation protocol.

Chapter 8 is the final empirical chapter. This third study investigates whether preferences over health states held by members of the general public are well informed. Participants express their beliefs about life in several health states and their beliefs are compared to the experience of patients. Data on patient experience was measured using secondary data.

The final chapter, Chapter 9, provides an overall integrated discussion of this thesis. It summarises the key findings and contributions of this thesis. The implications of the findings for methods for valuing health are discussed and recommendations for further research are made.

1.6 Conclusion

NICE regularly implements cost-effectiveness analysis to determine whether to recommend the use of interventions to the NHS (NICE, 2013a). To implement the cost-effectiveness analysis the benefits of interventions are measured using the QALY (Drummond et al., 2005, p.173; NICE, 2013a). The calculation of the QALY requires quality weights (Drummond et al., 2005, p.174). These weights thus help determine which treatments the NHS provides.

Quality weights are generally estimated using choice-based methods (NICE, 2013a). There have been concerns about the validity of certain assumptions of choice-based methods, in particular the assumptions of complete preferences, informed preferences, and the treatment of preferences as tastes. These assumptions may not be valid in the health domain and the violations may have implications for the validity of choice-based methods for valuing health. The aim of this thesis is to investigate the appropriateness of these three assumptions.

The next chapter describes the rationale of the PhD study, first by describing choice-based methods used to value health and the role of the preference satisfaction theory of welfare in valuing health.

Chapter 2

Three assumptions of choice-based methods

To determine the cost-effectiveness of an intervention the cost and benefits of an intervention are compared to the alternatives (Drummond et al., 2005, p.8). The benefits of health interventions are measured by estimating its total expected QALYs. For example, in the EQ-5D-3L system the health state with some self-care problems, moderate pain or discomfort, and extreme anxiety or depression (i.e. the health state 12123) has a value of 0.187 (Brazier et al., 2007, p.202). An intervention that results in living in health state 12123 for a period of 10 years results in a total of 1.87 QALYs. Different methods exist to determine the value of a health state and choosing an appropriate method to value health states is an important question in health economics.

The aim of this chapter is to review the theoretical background for this PhD study, in particular aspects of normative economics as they relate to methods for valuing health. This chapter will argue that choice-based methods for valuing health are based on the preference satisfaction theory of welfare, a prominent theory in normative economics. It then describes the theory and three of its important assumptions. Challenges to these assumptions are presented. To assess whether the assumptions of the preference satisfaction theory of welfare are valid three important research directions are described. To begin, a brief overview of cost-effectiveness analysis is provided.

2.1 Economic evaluation and cost-effectiveness analysis

Resources are scarce but demand is not (Drummond et al., 2005, p.8). Publicly funded healthcare systems have various mechanisms available to allocate scarce resources. One

such mechanism is the use of economic evaluation. The aim of economic evaluation is to compare the cost and consequences of alternatives (Drummond et al., 2005, p.9). One type of economic evaluation is cost-effectiveness analysis (Drummond et al., 2005, p.137). Cost-effective analysis is undertaken in England, where NICE incorporates cost-effectiveness analyses into its considerations on which treatments to recommend for use in the NHS (NICE, 2013a).

The type of cost-effectiveness analysis conducted by NICE is referred to as a cost-utility or cost-per-QALY analysis (Drummond et al., 2005, pp.138-139). In this type of economic evaluation the benefits of an intervention are measured using the generic outcome of utility or the QALY (Drummond et al., 2005, p.139). To encourage consistency in the approach to conducting cost-effectiveness analysis NICE has developed a set of recommendations, which are outlined in the NICE "reference case" (NICE, 2013a). Among other things, the reference case recommends that QALYs be measured using the EQ-5D questionnaire (where the EQ-5D is appropriate) (NICE, 2013a). The EQ-5D is a questionnaire developed by the EuroQoL group (Rabin and de Charro, 2001). It contains two versions, both with the same five domains of mobility, self-care, usual activities, pain or discomfort, and anxiety or depression (Rabin and de Charro, 2001). The EQ-5D-3L contains three levels per domain (Rabin and de Charro, 2001) and the newly developed EQ-5D-5L contains five levels per domain (Herdman et al., 2011).

In cost-per-QALY analysis an incremental cost-effectiveness ratio (ICER) is calculated for an intervention to be assessed (Drummond et al., 2005, pp.38-39). The incremental aspect means that each intervention is compared to the next best alternative intervention. The ICER is then compared to a pre-determined threshold (Brazier et al., 2007, pp.276-277). Everything else being equal, a lower ICER increases the chance of

recommendation because it means that the QALYs gained were achieved at lower cost[1]. In cost-per-QALY analysis the denominator of the ICER is expressed QALYs (Brazier et al., 2007, pp.276-277). The QALY is thus the measure of benefits of health interventions used in the type of cost-effectiveness analysis conducted by NICE (Drummond et al., 2005, p.173; Brazier et al., 2007, pp.10-11).

2.1.1 The quality adjusted life years

One QALY represents one year in full health (Brazier et al., 2007, pp.10-11). The total expected QALYs provided by a treatment is a combination of the length of life and the quality associated with the health states resulting from that treatment (Drummond et al., 2005, p.173). To determine the total QALYs produced by an intervention length of life is multiplied by the quality weights. For example, if someone lived in full health for 10 years they would obtain 10 QALYS (10*1) but if someone lived in the EQ-5D-3L health state 12123 for 10 years they would only obtain 1.87 QALYs (0.187*10) (Brazier et al., 2007, p.202).

Length of life is simply expressed in number of life years. Assessment of quality of a health state is more difficult and requires the determination of 'quality weights' (Drummond et al., 2005, p.174). By agreement, the quality weights are anchored on a weight of one being equivalent to 'full health' and a weight of zero being a state equivalent to being dead (Drummond et al., 2005, p.174), with negative numbers indicating a health state is worse than being dead. To determine the quality weight beyond the anchors of zero and one, various methods are available. The methods recommended in the NICE reference case, and generally endorsed in the literature, are referred to as choice-based methods (Drummond et al., 2005, p.145; NICE, 2013a).

---

1. At least this is the case for interventions where both cost and effectiveness increase compared to the alternative.

2.1.2 Choice-based methods to measure the quality of health states

To determine the quality weights of a health state economists undertake valuation surveys asking members of the general public to engage in a stated preferences exercise. Stated preferences exercises are so called because individuals are asked to state their preferences, as opposed to revealing their preferences by means of action (Brazier et al., 2007, pp.13-14; Morris et al., 2007, pp.267-268). Economists generally cannot observe preferences over health states being revealed because there is usually no market available where consumers can purchase care or care may be subsidised at point of purchase (Brazier et al., 2007, pp.13-14; Morris et al., 2007, pp.267-268). Various stated preferences tasks exist such as the time trade off (TTO), the standard gamble (SG), or Discrete Choice Experiments (DCEs) (Torrance and Change, 1986; Brazier et al., 2007, pp.87-93).

Choice-based methods of valuing health generally ask participants to imagine living in ill health and to make choices between different profiles of health (Torrance and Change, 1986). In the TTO participants are asked to trade length of life for a quality of life, in the SG to trade risk of being dead for quality of life, in DCEs to trade one health profile for another (Brazier et al., 2007, pp.84-106). For example, in the TTO a participant may be asked how many years in full health is equivalent to 10 years in 12111. If s/he responds with eight, his/her valuation for the health state 12111 is 0.8 (8/10). The important similarity between these techniques for health state valuation is that they aim to measure or elicit preferences. Hence, preferences are central to the valuation of health states (Drummond et al., 2005, p.174).

Choice-based methods are methods such as the SG, the TTO, and DCEs. These methods are said to be choice-based because respondents are asked to make choices (Brazier et al., 1999a). Choice-based methods are preferred by NICE (NICE, 2013b) and some health economists because the act of choice involves a trade-off of the

opportunity cost of the foregone alternative (Brazier et al., 1999a). Considering opportunity costs grounds these methods in established consumer choice theory and welfare economics (Green et al., 2000). Methods such as SG, TTO, and DCE have been linked to existing economic theory: the SG is linked to expected utility theory (Green et al., 2000), the TTO to Hicks utility theory (Buckingham and Devlin, 2006), and DCEs to the random utility model (Lancsar and Louviere, 2008; Louviere et al., 2010). The use of choice-based methods excludes one method that has been frequently used to value health, the VAS. The VAS is said not to be choice-based because participants are not forced to choose and hence do not face an opportunity cost (Brazier et al., 1999a). The lack of consideration of opportunity cost and trade-offs disconnects the VAS from wider economic theory, and indeed VAS is said to have its theoretical foundation in psychology (Parkin and Devlin, 2006).

The phrase choice-based and the superiority of choice-based methods have been debated in the literature (Parkin and Devlin, 2006). It has been argued that the VAS does involve choice because respondents have to make a choice of where to put the health state on a scale, involving a choice between other potential states on that scale and the anchor points (Parkin and Devlin, 2006). The hesitation to use the VAS may not necessarily be based on the importance of choice or trade-offs, but rather on more practical concerns (Torrance et al., 2001; Brazier and McCabe, 2007). The VAS: (a) produces different values to the SG and TTO (Torrance et al., 2001), (b) elicits different thought processes, which may mean participants ignore duration and their responses cannot be interpreted as preferences over health states (Robinson et al., 1997), (c) includes biases that appear more problematic than the biases present in the TTO and SG (Torrance et al., 2001; Brazier and McCabe, 2007).

Despite the reservation about the phrase choice-based, it is used as standard usage in the literature. Investigating the assumptions relating these 'choice-based' methods is

important, regardless of their true or false superiority compared to non-choice-based methods.

The reason for the importance of preferences for valuing health can be traced back to the field of normative economics. Normative economics is the branch of economics that deals with making judgements about whether certain states of the world are better or worse than others (Morris et al., 2007, p.210; Brouwer et al., 2008). Judging states of the world is important in informing policy making, for example policy making on health care spending (Morris et al., 2007, p.210). The evaluation of the benefits of health care spending thus relies on normative economics to determine what the benefits of a health intervention are and to determine how to measure and value those benefits[2] (Morris et al., 2007, p.210).

A prominent theory within normative economics is welfarism (Morris et al., 2007, p.210). A prominent theory of welfarism is the preference satisfaction theory of welfare (Hausman, 2006; Brouwer et al., 2008; Beckerman, 2011, p.39). The preference satisfaction theory of welfare specifies the 'ultimate goal' of all actions and the constituent of that 'ultimate goal' (Brouwer et al., 2008). By specifying what is 'good' and how to measure it, the theory acts as a guide on how to measure and value the benefits of health interventions. The next section reviews the preference satisfaction theory.

2.2 Preference Satisfaction theory of welfare

The preference satisfaction theory argues that welfare is the ultimate goal and that the satisfaction of an individual's preferences constitutes welfare (Brouwer et al., 2008; Hausman and McPherson, 2009; Beckerman, 2011, p.37). Individuals are said to have

2. This is not to say that normative economics is the only way of judging states of the world, but it is the method used in economics.

preference orderings of states of the world, i.e. they rank states of the world as more or less desirable. Their preferences are more satisfied as higher ranking states are reached[3] (Brouwer et al., 2008; Hausman and McPherson, 2009). Higher levels of preference satisfaction are better because they result in more welfare. Brouwer et al. (2008) describe this as the "the utility principle…individuals rationally maximize their welfare by ordering options and choosing the preferred option." This theory implies that the individual is the best judge of his/her welfare because the individual's preferences are what determine his/her welfare (Morris et al., 2007, p.211; Brouwer et al., 2008). These preference orderings can be represented by a utility function (Brouwer et al., 2008), although this requires a few axioms such as completeness, transitivity, and independence (Beckerman, 2011, p.39; Hausman, 2012b, p.13).

Choice-based methods of valuing health are used to implement the preference satisfaction theory to value the benefits of health interventions. Choice-based methods aim to measure improvements in welfare. The theory suggests that improvements in welfare can be measured by consulting an individual's preference orderings. Preference orderings are represented by an individual's utility function and therefore the health economist aims to measure an individual's utility function. Valuing health is thus a matter of eliciting preferences orderings over health states, using choice-based tasks such as the TTO or SG.[4]

In many ways the health economics field has expanded beyond welfarism, but the preference-satisfaction theory of welfare remains a guiding theory on how to value health. The ideas of extra-welfarism and non-welfarism are arguably as well accepted as welfarism (Culyer, 1989). Extra-welfarism and non-welfarism challenge some of the

3. This does not mean that the psychological state of feeling more satisfied has been achieved but that the state of the world is more like the state that is preferred.
4. This thesis does not discuss what happens after these valuations are elicited, for example how they should be aggregated and what this implies for a social welfare function because these issues are not the main focus of this thesis.

restrictions of welfarism (Brazier et al., 2007, p.51). For example, in an extra-welfarist framework it is possible to use outcomes other than utility and valuations by those other than the affected individuals (Brouwer et al., 2008). Certain aspects of conventional methods may in fact be more extra-welfarist than welfarist (Morris et al., 2007, p.236). For example, health state values are aggregated to form community values and it is not clear whether this is acceptable under welfarism (Torrance and Feeny, 1989; Gandjour, 2010). Furthermore, a welfarist may prefer the use of money as an outcome measure (Torrance, 2006). Despite the differences between welfarism and extra-welfarism, the acceptance of extra-welfarism has not changed the methods used to value health states. Although extra-welfarism may make a difference to cost-effectiveness methods, especially when it comes to interpersonal comparisons of QALY values or the application of QALY weights, health economists still value health by measuring preferences (Brouwer et al., 2008). In practice, extra-welfarism has made little difference to the focus on preferences when valuing health. Validating the assumptions of the preference satisfaction theory therefore remains important, even in a non-welfarist framework.

This thesis investigates the validity of the assumptions of choice-based methods, but does not assess external validity. A valid measure is one that measures the concept it is meant to measure (Bryman, 2012, p.47) and a valid measure of health state preferences is a measure that measures health state preferences. Traditionally, the test of 'external validity' of stated preferences has been to compare stated preferences to revealed preference (Louviere, 1988; de Bekker-Grob et al., 2012). If the two preferences are similar, then stated preferences reflect real preferences and reflect what individuals are likely to do in the real world (although this still ignores potential uninformed revealed preferences). An external test of validity is not available in the health care market, because individuals cannot be observed trading between health

states (Brazier et al., 2007, pp.13-14; Morris et al., 2007, pp.267-268). There may thus be no way to show that the preference elicitation tasks are externally valid. The focus of validity in this thesis is testing the internal assumptions of preference elicitation tasks. The appropriateness of the assumptions of the methods for valuing health would provide some evidence for validity, but may not be sufficient to declare that the elicited preferences are completely valid.

2.3 Assumptions of the preference satisfaction theory

The preference satisfaction theory of welfare rests on several assumptions. This thesis investigates three specific assumptions of the preference satisfaction theory of welfare: the assumption of complete preferences, the assumption of informed preferences, and the assumption that preferences are matters of taste. Although other assumptions are necessary or desirable (Drummond et al., 2005, p.142; Lancsar and Louviere, 2006), these three assumptions are, arguably, important and insufficiently studied in health economics. The following three sections present and discuss these three assumptions. For each assumption an alternative viewpoint is proposed and a method for testing the assumption is suggested.

2.3.1 Complete preferences

One of the basic axioms concerning preferences is that of completeness (Morris et al., 2007, p.24; Beckerman, 2011, p.36). Completeness means that an individual is able to establish a preference ordering between any two alternatives, such as two commodity bundles (Morris et al., 2007, p.24; Beckerman, 2011, p.36). When presented with two alternatives, an individual will be able to state that they prefer one to the other or is indifferent between the two (Hausman, 2012b, p.14). When asked for a choice between health state 11112 and 11121 an individual should be able to provide a preference

ordering of the two health states. Importantly, it is generally implied that the preference ordering exists independent of, or prior to, the elicitation task. Fischhoff (1991) argues that the assumption of pre-existing preferences implies a "philosophy of articulated values". Researchers who believe in the philosophy of articulated values believe that individuals are able to articulate and express preferences on a wide range of topics (Fischhoff, 1991). Furthermore, the preferences expressed are thoughtful and precise (Fischhoff, 1991).

The assumption of complete preferences suggests that an economist interested in an individual's preferences needs to focus on developing methods to elicit pre-existing preferences (Payne et al., 1999). When designing the task, the economist may believe, "if we've got questions, then they've got answers" (Fischhoff, 1991). This belief has implications for the design of preference elicitations tasks in health economics. Relatively quick and one-off interviews are the norm, with little or no opportunity for reflection during the task (Brazier et al., 2007, p.110). Participants are often not given the opportunity to change their preferences during health state valuation tasks, such as those based on the Measurement and Valuation of Health (MVH) group protocol (Gudex, 1994; Dolan et al., 1996a).

The focus on eliciting pre-existing preferences does not mean that the design of the elicitation task is simple. The economist needs to ensure that the right preferences are elicited and therefore the questions posed, or the alternatives under consideration, need to be defined very precisely and understood as intended (Fischhoff, 1991; Payne et al., 1999). In addition, the respondent must have the motivation to respond truthfully (Fischhoff, 1991; Payne et al., 1999). These concerns are apparent in discussions about methods used to value health. Methods such as TTO, SG, or VAS, are judged by how well they are believed to elicit preferences (Brazier et al., 2007, p.105) and health economists are concerned about whether patients will behave strategically by providing

untruthful responses to serve their personal interests (Brazier et al., 2007, p.115). There is less worry about the question of whether the preferences that are to be elicited actually exist. Research investigating the completeness of preferences has taken place, but many of the choice-based methods of valuing health can be characterised by an implied belief in completeness and articulated values.

In reality, preference orderings may not be complete. Even the originators of the expected utility theory questioned the assumption that preferences are complete:

> *This procedure for a numerical measurement of the utilities of the individual depends, of course, upon the hypothesis of completeness in the system of individual preferences. It is conceivable - and may even in a way be more realistic - to allow for cases where the individual is neither able to state which of two alternatives he prefers nor that they are equally desirable...How real this possibility is...seems to be an extremely interesting question, but it is a question of fact. It certainly deserves further study. (Neumann and Morgenstern, 2007, p. 19)*

A sizable literature in psychology indeed suggests that preferences are generally not complete and merely elicited, but rather constructed during the elicitation task (Payne et al., 1992; Slovic, 1995). The evidence for this position is the large literature showing that preferences are not description invariant, meaning that an equivalent but different description of options can influence preferences (Payne et al., 1992). For example, describing the exact same outcome as number of lives saved or number of lives lost results in different preference orderings (Payne et al., 1992). Preferences are also not procedurally invariant, meaning that equivalent but different methods of elicitation can influence preferences (Payne et al., 1992). For example, choosing between program A and B with different benefits and costs can be done by choice or matching methods. In choice methods individuals are asked to choose between program A and B. In matching methods, an individual is asked how much the cost of program B should be before they are indifferent between the two. Preferences elicited using choice and matching methods have been shown to be inconsistent (Payne et al., 1992). The lack of

descriptive and procedural invariances can be caused by various mechanisms, for example individuals may use different decision strategies in different situations, may simplify complex tasks, and may be uncertain about their values (Payne et al., 1992). Therefore, normatively irrelevant differences in description and procedures cause systematic differences in preferences (Payne et al., 1992; Slovic, 1995). The existing evidence suggests that individuals may not have existing preference orderings for all alternatives (Fischhoff, 1991). Individuals may have preferences over alternatives in only a few domains but may have values "on more general and fundamental issues" (Fischhoff, 1991).

Preferences over health states may better conform to the assumption of preference construction than of complete preferences.[5] Individuals are likely to have well-formed preferences on topics which are familiar, consequential, uncontroversial and simple (Fischhoff, 1991; Slovic, 1995). If the topic is complex and unfamiliar, preference elicitation tasks and are likely to contain a fair amount of preference construction unless individuals have had a chance to evaluate the alternatives previously (Fischhoff, 1991; Slovic, 1995). It can be argued that health state valuation is a complex and unfamiliar task. The task is complex because participants are asked to consider many (between five to eight) different aspects of health in addition to probabilities and survival duration (Brazier et al., 2007, p.109). The task is unfamiliar because individuals do not generally face choices in health state valuation tasks in their daily lives (Fischhoff, 1991; Hausman, 2006). The nature of health state valuation tasks suggest that preferences over health states may be constructed rather than complete.

Empirical work directly assessing completeness over health state preferences is difficult and has been inconclusive. Shiell et al. (2000) attempted to test completeness

5. The choice between either complete or incomplete preferences is a false dichotomy. Preferences could be somewhat incomplete or one could maintain the philosophy of partial perspectives (Fischhoff, 1991). The point is that preferences may not be entirely complete.

by measuring whether preferences over multiple administrations of the SG were stable. Participants who changed their responses were assumed to have incomplete preferences. Overall, about 30% of their participants were said to have incomplete preferences. Ryan and San Miguel (2003) attempted to assess completeness of preferences over health care goods using a DCE task. Preferences over supermarket goods, dentist consultation, and bowel screening were compared. Completeness was investigated by assessing whether participants repeated the same preference orderings when a question was repeated in the same questionnaire. It was expected that preferences for supermarket goods are more complete than for the two health care goods and that therefore more respondents would provide the same preferences for the supermarket good when the question was repeated. Contrary to this hypothesis, Ryan and San Miguel (2003) found that there was little difference between completeness of the supermarket goods and the health care interventions. Therefore, preferences over health care interventions could were not found to be incomplete. Whether using stability of preferences can test completeness has been debated. Oliver (2000) argues that completeness merely requires participants to state a preference ordering and that individuals without an existing preference ordering can express preference to please the researcher. This "makes completeness difficult to test, and may be a reason why it has been neglected in the literature" (Oliver, 2000). Hougaard et al. (2012) argue that random error or random choices and incomplete preferences will be difficult to distinguish. There is also the possibility that individuals simply recall their previous responses, which will tend to make their preferences appear as complete (Ryan et al., 2009). It is therefore difficult to directly test completeness of preferences.

An alternative method of testing completeness is to investigate the assumption indirectly. Health state preferences can be elicited using methods that don't assume completeness and those preferences can be compared with methods that do assume

completeness. This indirect investigation can reveal how practically important the completeness and articulated values assumptions are for health state valuation. As Fischhoff (1991) argues, "the magnitude of an effect provides the critical test of whether it is worthy of notice. Unless it can be shown to make a difference, who cares?" It may be thus more practical to first consider how health state valuation tasks would change if the preference construction view were accepted.

If preferences are constructed during the task, health state valuation tasks should be designed accordingly. Two types of suggestions are made in the literature. First, several authors have suggested that participants can be allowed to reflect or deliberate on their preferences (Sen, 1987, pp.14-24; Fischhoff, 1991; Dolan, 1997b; Hausman, 2010). Although reflection and deliberation are often used interchangeably they refer to different things (Goodin and Niemeyer, 2003). During reflection participants can examine their preferences individually (Sen, 1987, pp.14-24). Reflection is defined as "the action or process of thinking carefully or deeply about a particular subject, typically involving influence from one's past life and experiences" (Oxford University Press, 2015). Participants could also be allowed to engage in deliberation with other members of the public (Dolan, 1997b; Hausman, 2010). Deliberation is defined as "a discussion and consideration by a group of persons" (Merriam-Webster, 2015). If completeness is not assumed, then the health state valuation task could provide participants with an opportunity for reflection and deliberation.

The second suggestion is to guide participants to develop their preferences based on more general values (Fischhoff, 1991). Importantly, Fischhoff (1991) never specifies what general values are and they may depend on the domain over which preference are elicited. Participants can be guided by the use of a structured decision making process (Fischhoff, 1991; Gregory and Slovic, 1997). For example, in the context of environmental valuation the use of Multi-Attribute Utility Theory (MAUT) framework

is suggested (Gregory and Slovic, 1997). MAUT is a specific form of Multiple Criteria Decision Analysis (MCDA)(Huang et al., 2011), and more specifically of the "value measurement model" type of MCDA (Belton and Stewart, 2001, p.9).

MCDA is a systematic process that assists individuals in choosing between options when there are conflicting criteria (Belton and Stewart, 2001, p.2). The details of this process are explained in Chapter 6, but are explained briefly here. The "value measurement model" type of MCDA process involves determining options (or choices), criteria by which the options can be judged, scoring the options against the criteria, weighing the criteria, and aggregating the scores and weights. Criteria are the factors that make one option better or worse than another (Belton and Stewart, 2001, p.1). For example, price is usually a criterion when purchasing a car. Once criteria are determined (i.e. 'price and safety matter when purchasing a car'), participants score each option on each criterion. Scoring thus assesses how well each option performs on each criterion (i.e. 'how well do the various cars compare on the criteria of price and safety'). Participants then weigh each criterion ('how valuable are price and safety'). The weights and the scores can be combined in a mathematical formula to develop a score and ranking for each option. The fundamental aspect of the MCDA process is that participants are guided through the process of developing their preferences by considering more basic aspects of their values and their options.[6]

Two suggestions are made in the literature on how to deal with incomplete preferences: (i) reflection and deliberation, and (ii) MCDA. MCDA can be considered as a type of reflection, as the process of MCDA helps individuals examine their preferences (Belton and Stewart, 2001, pp.3-5). Thus, reflection, with the use of MCDA, and deliberation are potential tools for designing a health state valuation task

6. Although MAUT has been applied in health economics to value the HUI descriptive system, it has not been applied in the context of comparing MAUT based preferences to other preference based methods.

that do not assume completeness, but rather assume that preferences are constructed. The comparison of preferences elicited with and without reflection and deliberation can serve as an indirect test of completeness.

2.3.2 Informed preferences

Economists have developed criteria to determine whether preference orderings are rational (Beckerman, 2011, pp.39-40). The rationality of preferences is usually defined as a combination consistency and transitivity (Sen, 2002, p.ch3; Morris et al., 2007, p.24; Beckerman, 2011, p.39). Consistency means that if alternative A is preferred to alternative B, the preference ordering will remain unless a relevant change in the environment takes place. Transitivity means that that if health state 11112 is preferred over 21111 and health state 21111 is preferred over 31111 then health state 11112 should be preferred over 31111. The two criteria are both matters of 'internal consistency' or 'internal coherence' because they do not require information of anything beyond the preference orderings themselves (Kahneman, 1994; Sen, 2002, p.ch3). Rationality of preferences is thus generally limited to criteria of internal coherence.

Sen (2002, p.19) has argued that preferences ought to be more than internally coherent. Internal coherence is not a sufficient condition for assessing the 'quality' of preferences because "a person can be consistently moronic in his choices" (Sen, 2002, p.20). It is possible to consider a preference ordering internally coherent, and thus rational, yet 'wrong'. Preference orderings can be rational but 'wrong' because preferences can be evaluated based on external criteria, such as the goals of the individual (Bhattacharyya et al., 2011). For example, an individual's goal can be to maximise hedonic utility. It is possible to imagine an internally consistent and rational set of preferences that do not maximise hedonic utility. This can occur because an

individual can make an error in forecasting future tastes or evaluating past experiences (Kahneman, 1994).

Empirical evidence of rational but apparently arbitrary stated preferences has been reported. For example, Ariely et al. (2003) elicited willingness to pay amounts for several goods in a stated preference exercise. The goods could be ranked on quality and quantity. The authors found that participants expressed willingness to pay amounts that appeared to be generated by an internally coherent preference ordering. But the willingness to pay amounts were also related to arbitrary anchors such as a participant's social security number. The preference orderings were internally coherent because "people respond in a robust and sensible fashion to noticeable changes or differences in relevant variables" but also arbitrary because "responses occur around a base-level that is normatively arbitrary" (Ariely et al., 2003). In the context of health state valuation, one can argue that just because health state values appear internally coherent, does not mean they are not arbitrary. Some other criterion of beyond rationality may thus be required.

One example of a quality criterion for preferences beyond rationality is that of informed preferences. There have been arguments that, at least in welfare economics, informed preferences should be elicited (Hausman, 2006; Beckerman, 2011, p.38). Informed preferences are those an individual would have if s/he "had all the relevant information and made full use of this information" (Harsanyi, 1997). Similarly, in the health economics literature there have been arguments for the use of preferences from individuals who are fully informed (Gold et al., 1996, pp.99-100; Brazier et al., 2005). Yet there are reasons to believe that preferences over health states may not be informed (Hausman, 2006). For example, most participants are asked to value states for which they have little experience (Hausman, 2006; Brazier et al., 2007).

Preferences over health states should be rational but also informed. To determine whether preferences are informed one must determine whether the individual is informed about what they are valuing and assess the factual information that participants have during health state valuation tasks. Preferences over health states should thus be tested to see if they are informed.

2.3.3 Preferences as tastes

The last assumption discussed in this thesis is the tendency for some economists to treat preferences as matters of only taste (Hausman, 2011). Treating preferences as tastes is not technically an axiom or assumption of the preference satisfaction theory of welfare, but it is sometimes implicitly accepted (Hausman, 2011). This may be due to the belief that economists need not concern themselves with how preferences are formed (Hausman, 2012b, p.8), for example Stigler and Becker (1977) note:

> *On the traditional view, an explanation of economic phenomena that reaches a difference in tastes between people or times is the terminus of the argument: the problem is abandoned at this point to whoever studies and explains tastes (psychologists? anthropologists? phrenologists? sociobiologists?)*[7]

Whatever the reason, economists sometimes treat all preferences as matters of taste.

The difference between preferences and tastes are that tastes do not rely on beliefs, they are only matters of 'gut instincts' (Hausman, 2011). In contrast, some preferences depend on beliefs and are not simply matters of gut instinct (Hausman, 2011). Hausman (2006) argues for a distinction between two types of preferences. Basic-preferences are similar to tastes and do not depend on any beliefs about the world. Non-basic preferences may depend on tastes but also depend on beliefs about the world (Hausman, 2006).

---

7. Stigler and Becker (1977) later argue against this view: "on our preferred interpretation, one never reaches this impasse: the economist continues to search for differences in prices or incomes to explain any differences or changes in behaviour."

Preferences over ice cream can be an example of a basic-preference. The preference of chocolate over vanilla need not include beliefs or technical judgements and can be simply a matter of taste. An individual has the taste for vanilla and there is no scope for argument with a chocolate fanatic on pure matters of taste.[8] Preferences over medical interventions can be non-basic. An individual's preference for taking a certain drug is dependent on their beliefs about the benefits of that drug (Hausman, 2006) and perhaps some basic 'taste' for survival[9]. An individual's preferences over jobs may depend on the beliefs of the individual about the particular aspects of the jobs (how interesting the work is, how high the salary is, and how convenient the location is) combined with beliefs about how important those aspects are (Harsanyi, 1997).

Non-basic preferences should not be treated as matters of taste because non-basic preferences can be subjected to investigation and discussion (Harsanyi, 1997; Hausman, 2006). The maxim highlighted by Stigler and Becker (1977), 'De gustibus non est disputandum' ('regarding taste, there is no dispute') is applicable only to basic-preferences. One cannot argue about matters of gut instinct because no reasons can be provided for them. This is not the case for non-basic preferences. For example, the beliefs underlying the preference of taking one drug over another can be investigated. Non-basic preferences involve beliefs and beliefs are open to argument (Hausman, 2006; Hausman, 2011).

Treating non-basic preferences as basic means that non-basic preferences will not be fully understood because the underlying beliefs are not investigated. Better understanding of preferences can have two benefits. It enables better development of

8. It is possible that preferences over ice-cream are non-basic. They may depend on tastes as well as beliefs, for example whether one flavour is healthier than another. Hausman (2006) suggests that there may be no basic-preferences at all.

9. There are parallels here to the concept of supplier-induced demand studied in health economics. In the interaction between the patient and physician, the goals or preferences of the principal-patient need to be combined with the technical judgements, or beliefs, of the agent-physician in order to make choices. the principal lacks the technical judgements which leads to the principal-agent problem (and the potential for supplier induced demand) (Morris et al., 2007, p.49).

elicitation methods and allows the preferences to be critically assessed. It can have advantages for developing methods to value health because those methods have to be sensitive to the complexity of underlying beliefs. It can have advantages for critically assessing the preferences over health states because the underlying beliefs can be assessed. Therefore, investigating non-basic preferences can help to better understand those preferences and to potentially adapt valuation tasks accordingly.

Hausman (2006) argues that preferences for health states are non-basic because they must involve a series of judgements and beliefs. Individuals have to make judgements about the health states they are valuing in the context of particular environments they are living in (Hausman, 2006). It is recognised that health state valuation are cognitively complex and require experience, information, and judgements (Brazier et al., 2007, pp.109-118). It is thus likely that preferences over health states are non-basic. The beliefs underlying preferences over health states can be investigated to whether they are purely matters of taste or not.

2.3.4 Summary and key research directions

In the previous section three assumptions of the theory of preference satisfaction of welfare were discussed, namely that preference are assumed to be complete, informed, and purely matters of taste. These three assumptions mean that economists entirely following the welfare economics tradition should not investigate or judge preferences as such, but should focus on asking exactly the right question in order to elicit the right preference ordering. As Brouwer et al. (2008) state "...welfarists do not usually enquire as to how these preferences may have come about nor do they judge them on grounds of decency, ethics, aesthetics or any other ground". Preferences elicited using tasks with one-off individual interviews are viewed as a satisfactory method of valuing health

because they measure preferences and conform to the assumptions of the preference satisfaction theory of welfare.

This chapter presented challenges to those three assumptions. First, it was argued that preferences may not be complete. The assumption of preference construction may be more applicable to preferences over health states. It was suggested that reflection and deliberation could be investigated as methods to guide individuals to construct their preferences. Second, it was argued that rational preferences may be inadequate in welfare analysis. Individuals should also have informed preferences. Finally, it was argued that not all preferences are matters of taste and that some preferences depend on beliefs. Preferences over health states may be non-basic and may include beliefs about the world. If so, researchers should investigate and discover those beliefs.

Further research is required to discover whether the alternative viewpoints are more applicable to preferences over health states. The three challenges to the assumptions presented in this chapter imply the need for research on:

1) the role of reflection and deliberation (including MCDA) in health state valuation
2) the assessment of whether preferences over health states are informed
3) the examination of the judgements, cognitive processes, and beliefs underlying preferences over health states

2.4 Conclusion

In this chapter it was argued that the preference satisfaction theory of welfare underlies the design of choice-based methods, such as the TTO and SG, for valuing health. Three assumptions of the preference satisfaction theory of welfare were discussed. Existing methods of valuing health may not be valid if the assumptions are violated. Invalid methods could potentially result in the wrong policy decisions on the important issue of

resource allocation amongst health care interventions. Therefore, it is important to investigate the validity of choice-based methods for valuing health by investigating the appropriateness of the assumptions. To investigate the assumptions three key research areas were suggested.

The next chapter reports a literature review on the three key research areas. It describes and discusses the existing literature, and highlights the research gaps of the literature.

Chapter 3

A review of the literature

In this chapter a literature review is conducted to review existing knowledge about the three key research areas of this PhD:

1) the role of reflection and deliberation (including MCDA) in health state valuation
2) the assessment of whether preferences over health states are informed
3) the examination of the judgements, cognitive processes, and beliefs underlying preferences over health states

This chapter summarises the findings from the literature and highlights the gaps in existing knowledge.

3.1 Literature review method

This review is not a systematic review (Grant and Booth, 2009) but is conducted in a systematic fashion. The nature of this review is methodological. It is not a review with the aim of arriving at a precise point estimate based on a single study design and a narrow research area (Edwards et al., 1998, p.255), as is typical in the Cochrane type of systematic reviews (Grant and Booth, 2009). The abstract concepts and ideas associated with the topic and the methodological nature of this review make it likely that a conventional keyword based search will result in a large number of irrelevant references and therefore a systematic review is not necessarily appropriate (Edwards et al., 1998, pp.257-258; Dolan et al., 2005). Despite this, there is still a benefit to conducting the review in a systematic fashion.

Three methods are used in this literature review. The first method is "citation pearl growing" (Booth, 2008). In this method a set of key papers known to the PhD student and supervisors are identified. Key papers are papers that are relevant to the literature review and form the basis of the search strategy. The key papers are reviewed for keywords. These keywords are used as search phrases in a database search. The database results are filtered for relevant papers. If further relevant papers are identified then those papers are reviewed for new keywords. Any new keywords are added to an updated search strategy and the updated search is conducted. The updating process continues until no new papers and keywords are identified. A second method is "footnote chasing" where the references of identified papers are reviewed to see if they contain any relevant papers (Booth, 2008). The third method used is "citation searching" where any papers that have cited already identified papers are reviewed (Booth, 2008). Any new keywords from papers identified using footnote chasing and citation searching are also included in the search strategy. The approach using the three methods is conducted for several iterations until no new relevant papers are identified. Finally, speculative searching is conducted to find any unpublished papers. The entire process of literature searching is shown in Figure 3.1.

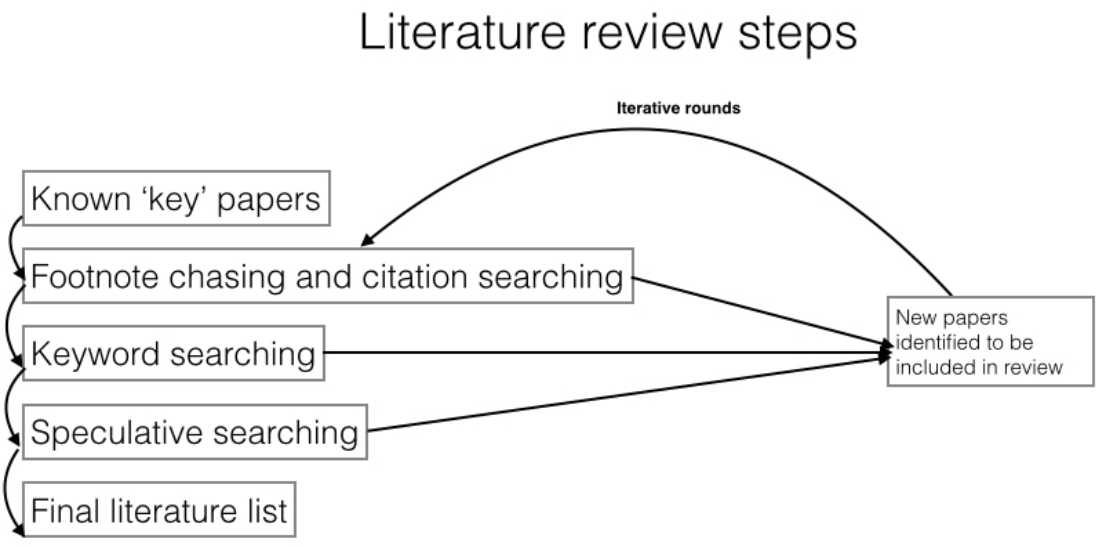


Figure 3.1 Steps used for the literature search

Google Scholar (Google, 2015b) and Web of Science (Thomson Reuters, 2015) were used to find papers that cited the identified papers. Speculative searching was conducted using Google search engine (Google, 2015a).

To conduct the keyword search two databases were searched, SCOPUS (Elsevier, 2015) and OVID (Ovid Technologies, 2015). OVID was used to search Econlit, MEDLINE, and PsycINFO. The SCOPUS databases contain a large and broad range of the literature (Vieira and Gomes, 2009) and the databases accessed with OVID have a particular focus on economics, medicine, and psychology literature, which is important because relevant research in health economics can span all those fields. The search start date was variable depending on the database and is reported in Appendix 1, but the last search date was June 2015 (the original review conducted before the data collection for the three studies was updated before submission of this PhD study). The full search strategy is reported in Appendix 1. The search included titles, abstracts, and keywords.

Two inclusion criteria were used. Studies that valued health states using often used methods for valuing health states (e.g. VAS, TTO, DCEs) were included. Non-English papers were excluded.

In the following sections details of the search strategy are given and the relevant papers are discussed. Next, the literature is summarised in a narrative format, the gaps in the literature are discussed, and the implications of the gaps for this thesis are highlighted.

3.2 Overview of papers found

The literature review process was conducted for the three research areas. Papers on reflection and deliberation (question one) and on judgments underlying health state values (question three) contained similar keywords. Therefore, the search strategy for these two questions was combined. For the review on whether preferences are informed no relevant papers were identified. A speculative search undertaken with Google search engine (Google, 2015a) did not identify any relevant papers. Therefore no discussion of that review is presented in this chapter.

The literature search first started with identifying known papers, searching their references, and searching papers that cited them. The PhD student and supervisory team were aware of three papers about the judgments underlying health state values (Robinson et al., 1997; Baker and Robinson, 2004; Mulhern et al., 2012). One relevant paper was found from the references (Spencer, 2000) and four were identified because they cited one of the three original papers (Devlin et al., 2004; van Osch and Stiggelbout, 2005; Oliver, 2007; Osch and Stiggelbout, 2007). The PhD student and supervisory team were aware of three papers on reflection and deliberation (Shiell et al., 2003; Stein et al., 2006; McTaggart-Cowan et al., 2011). Two additional papers were found because they are cited in one of the three key papers (Damschroder et al., 2005b; Damschroder et al., 2008). Two papers were found because they cited one of the key papers (McIntosh et al., 2007; Robinson and Bryan, 2013). Additional rounds of footnote chasing and citation searching did not find any more relevant papers.

Next, all papers were reviewed for keywords and these keywords were used in a database search. Several iterations of searching were undertaken as more papers were found. In the final search contained three types of keywords:

- Keywords about valuation (e.g. 'preferences' or 'utility')

- Keywords about health states or methods of valuation (e.g. 'health state' or 'time trade-off')
- Keywords about reflection or deliberation (e.g. 'group valuation' or 'reflection') or about investigating participants' thought process (e.g. 'qualitative' or 'reasoning')

A total of 230 unique papers were identified in the two databases. Most were removed because they were not relevant. A total of 179 were removed after reading the title and 29 removed after reading the abstract. One paper was removed after reading the full paper because the reported results were so brief that they were not useful for this thesis (Teixeira and Schackman, 2008). Another 15 were already found because they were either one of the key papers, were listed in the references of included papers, or had referenced included papers. The rest (six papers) were included. One paper about the use of reflection was found from speculative searching. Footnote chasing and citation searching was conducted on the final six papers but no additional papers were found. From the total of 22 papers, 10 discussed the effect of reflection and deliberation on health state and preferences 12 were about understanding the judgements underlying health state values.

The next sections present a summary of the literature and reports on the gaps in the literature with relation to the three main objectives of this thesis.

3.3 What is the effect of reflection and deliberation on health state preferences?

The first research question in this review is on attempts to investigate the effect of reflection and deliberation in health state valuation. Various methods have been used in the literature to encourage participants to reflect or deliberate before valuing health

states. In general, the literature can be divided into two types of methods: group-based and individual-based methods. In individual-based methods participants engage only in a solitary exercise. Group-based methods gather participants in a group and ask them to engage in some form of exercise with other participants. Group-based methods may include aspects of individual reflection as well. A list of the 10 included papers, their valuation method, the design of the reflection or deliberation method, sample size, findings, and source is shown in Table 3.1. The next sections present a summary of the findings and discuss the gaps in the literature.

Authors	Method	Design	Sample size	Relevant findings	Source paper
Shiell et al. (2003)	SG	Individual - Reflection exercise	92	No aggregate level change	Key paper
Peacock et al. (2004)	TTO	Individual - Reflection and deliberation exercise	365	No aggregate level change	Speculative searching
Damschroder et al. (2005b)	PTO	Individual - Adaptation exercise	359	Aggregate level change, values increased, more so for severe states	Cited in McTaggart-Cowan et al. (2011)
Damschroder et al. (2008)	TTO, SG	Individual - Adaptation exercise	1117	No aggregate level change	Cited in McTaggart-Cowan et al. (2011)
McTaggart-Cowan et al. (2011)	TTO	Individual - Adaptation exercise	90	Aggregate level change, values increased, more so for severe states	Key paper
Krabbe et al. (1996)	SG, TTO	Group - Interactive voting	104	Aggregate level change, values decreased, more so for severe states	Keyword searching
Stein et al. (2006)	SG	Group - Deliberation	15	No aggregate level change	Key paper
Akunne et al. (2006)	VAS	Group - Deliberation requiring consensus	70	Aggregate level change, values decreased	Keyword searching
McIntosh et al. (2007)	SG	Group - Deliberation	193	Aggregate level change, values decreased, more so for severe states	Cites Stein et al. (2006)
Robinson and Bryan (2013)	PTO	Group - Deliberation	50	Aggregate level change, values decreased	Cites Stein et al. (2006)

Table 3.1 Overview of papers included in review on reflection and group methods to value health

3.3.1 Individual-based methods

Five papers were found with individual-based exercises aimed to encourage participants to engage in reflection beyond what is normal in health state valuation exercises. Shiell et al. (2003) was the first to investigate the effect of reflection on health state preferences. A total of 92 participants valued health states three times over five weeks using the SG and TTO. Participants were divided into an intervention and a control group. After the first valuation, participants in the intervention group were provided with a booklet with their responses to the valuation task. The booklet provided an explanation of the valuation tasks and discussed the implications of the participants' responses for which aspects of health they valued most. The control group were only asked to return for the next interview. Valuations for the intervention and control group, over three occasions, were compared. The findings indicate that there was no difference between the mean values of the intervention and control group. The authors did not report the health state values. Shiell et al. (2003) concluded that the valuation techniques are adequate for the purpose of economic evaluation because aggregate level mean values are used in economic evaluations.

Peacock et al. (2004) tested whether engaging in a reflection exercise changed individuals' preferences over health states. Participants valued 11 health states, described using the Assessment of Quality of Life (AQoL) questionnaire, twice using the TTO. After the first valuation, respondents were randomly assigned to a control or intervention group. There were 143 participants in the intervention group and 223 in the control group. The intervention group was asked to engage in a reflection exercise at home after the first interview. The reflection task consisted of tasks such as completing the AQoL and a TTO task for the worst health they had ever experienced. It also asked participants to complete a TTO using the worst health state of the AQoL. These tasks were meant to encourage reflection and also familiarity with the

questionnaire. Respondents were also asked to discuss the valuation task with a family member or friend but whether and how they actually did this was not verified by the authors. Health state values were re-elicited after two to three weeks for both groups. Both parametric and non-parametric tests were used to compare health state values of the intervention and control group elicited during the second interview. The TTO valuations between the first and second interview were found to be statistically significantly different for some health states. But importantly, the second TTO valuations of the intervention and control group were found to be not statistically significantly different. The mean of the absolute difference between the intervention and control group over seven health states was 0.02. Mean standard deviation of the health state values for both groups was 0.31. The authors conclude that the intervention may have encouraged reflection, but had no effect on mean health state values beyond the effect of a repeat administration.

Damschroder et al. (2005b) asked 359 people to complete an online person trade-off (PTO) elicitation for two health states describing chronic pre-existing and new onset paraplegia. Half of the participants were randomly assigned to complete an 'adaptation exercise' meant to encourage respondents to reflect on their ability to emotionally adapt to negative circumstances. The adaptation exercise increased the aggregate value of a person's life in the chronic state from 98% to 100% of a person's life in full health. For the new onset state it increased values from 10% to 98%. Both changes were statistically significant. The increase in value suggests that asking participants to do an adaptation exercise increases the value participants place on saving the lives of people in ill health. Damschroder et al. (2005b) conclude that if members of the general public are not encouraged to think about their ability to adapt, their health state preferences may be flawed, even when using other valuation methods such as SG or TTO.

Damschroder et al. (2008) adapted the study by Damschroder et al. (2005b) to the TTO and SG. A total of 1,117 respondents valued four chronic health conditions using either the TTO or the SG. Before the valuation task half the respondents engaged in an adaptation exercise, similar to the previous study (Damschroder et al., 2005b), which was meant to get participants to reflect on their ability to adapt emotionally to difficult situations. The mean values of all four heath states were similar across both the intervention and control group, across both the TTO and SG and not statistically significant different. The mean of the absolute differences between control and intervention group was 0.02. The mean of the standard deviation of the samples was 0.28. Damschroder et al. (2008) argue that the adaptation exercise was able to get individuals to consider their ability to adapt but that this did not change their preferences. The authors suggest that one reason for the different result to Damschroder et al. (2005b) was the use of TTO and SG instead of the PTO, though why there would be a difference between the health state tasks could not be answered based on this study.

McTaggart-Cowan et al. (2011) further investigated the possibility of considering adaptation during health state valuation. A total of 90 participants valued three rheumatoid arthritis states before and after an adaptation exercise using the TTO. This exercise consisted of listening to recordings of patients discussing how they adapted to living with ill health. The adaptation exercise resulted in a statistically significant increase in the values for the states: 0.81 (SD 0.25) to 0.87 (0.22) for mild rheumatoid arthritis, 0.64 (0.32) to 0.70 (0.34) for moderate, and 0.25 (0.48) to 0.42 (0.50) for severe arthritis. Thus, increases were 0.06 for mild and moderate rheumatoid arthritis and 0.17 for severe arthritis. McTaggart-Cowan et al. (2011) conclude that an adaptation exercise did encourage individuals to increase their initial health state valuations. The results are contrary to the previous study (Damschroder et al., 2008), the authors suggest this

could be because their adaptation exercise was health specific while that used by Damschroder et al. (2008) was not.

From the five studies including some individual reflection exercise two found statistically significant changes at the aggregate level, with values increasing in both studies. Three studies did not show statistically significant changes.

3.3.2 Group-based methods

Five studies used a group-based exercise meant to encourage deliberation. Krabbe et al. (1996) compared average individual valuations with group valuations. The sample size was 104 and all participants completed an individual and group-based SG and TTO task. The two valuations took place within a 10 day interval. Group-based valuations were elicited using a modified version of the TTO and SG using an interactive voting system. Individuals were given pairwise choices. The pairwise choices were modified until the point at which both choices were selected by approximately 50% of the participants. This point was used as the indifference point of the group. For example, for the SG participants may have been presented with choice A of EQ-5D-3L state 12212 and choice B of a 16% chance of being dead and 84% chance of being in full health. If this resulted in about 50% of participants choosing A and 50% B, the value for state 12212 would be 0.84. The values were converted to a 0 to 100 scale.

The mean of the absolute difference in values between mean individual and group-based valuations was 11 points on a 0 to 100 scale. The differences were higher for severe states and much smaller for less severe states. The group-based valuations were lower than the mean of the individual valuations. It was not possible to statistically test differences between group and mean values because group values are only point estimates (i.e. the entire group produces only one value). The authors conclude that the individual and group-based values were different for the more severe health states.

Stein et al. (2006) investigated the completeness of preferences by assessing the effect of discussion on health state preferences. A total of 15 participants valued health states five times over six months. Preferences were elicited for 41 scenarios using the SG. After individually valuing a scenario the participants discussed the health states and were allowed to change their values. A total of 3.6% of the responses were changed. The mean of the absolute differences of the values before and after discussion was 0.02 and none of the changes for any of the health states were statistically significant. One limitation of this study was the small sample, which meant that statistical tests were underpowered. The study included semi-structured telephone interviews with participants after all group meetings had been conducted. The findings from the interviews indicated that participants valued the discussion for procedural reasons: greater reassurance, improved procedural performance, increased group cohesion, and satisfaction of curiosity. Stein et al. (2006) argue that assumption of completeness is "not unsafe" when using mean values.

Akunne et al. (2006) compared individual and group valuation of health states in several different communities in West Africa. Seventy participants, in seven groups, valued ten health states using a culturally adapted visual analogue scale. After first individually valuing health states, three or four respondents had to justify their scores to the group. After this participants had the choice to change their individual valuation. After the second round of individual valuation, the health states were discussed in the group and participants had to come to a consensus as a group on the value of each state. On average, the group valued the health states as more severe than the mean of individual valuations by 20%, on a scale from 0 to 1. The difference was statistically significant. The authors could not provide an explanation for the difference.

The Classification and Measurement System of Functional Health (CLAMES) was developed in Canada for use as a summary measure of population health (McIntosh et

al., 2007). During the valuation of the CLAMES, McIntosh et al. (2007) tested the effect of group deliberation on health state values. A total of 193 participants were split in groups of eight to 11 participants. Participants valued 12 states using the SG. After valuing each state, participants presented their scores on individual whiteboards and shared their reasoning. Participants then had the opportunity to change their preferences. There were statistically significant differences between the pre- and post-discussion mean values for five of the twelve states. Severe health states were more likely to change and to be valued lower after deliberation. The authors suggest this is perhaps because members of the public had little experience with the severe states. The mean absolute change in the value of the states that had statistically significant changes was 0.034. McIntosh et al. (2007) conclude that participants developed their preferences for some health states via discussion.

Robinson and Bryan (2013) explored the effect of deliberation on health state values using the PTO. In total 50 individuals took part in the study. Participants individually valued nine health states. After participants valued the health states, all valuations were displayed on a screen and the valuations and their implications were explored. After discussion, participants had the opportunity to change their values. A total of 28% of the valuations were changed and this resulted in lower values for eight states. The mean absolute change was 0.18. No statistical test of significance was reported. After the group meetings 20 participants were interviewed about their experience. Robinson and Bryan (2013) concluded that participants were not likely to change their valuations because of other participants' valuations but that discussing the implications of the valuations and group discussion around life with certain diseases did prompt changes. The authors conclude, "respondents may well have incomplete preferences".

From the five studies including group-based exercises four found statistically significant changes at the aggregate level. Values decreased in all four studies and were more likely to change for the more severe states. One study did not show statistically significant changes. In the next section both the individual-based and group-based studies are discussed and gaps in the literature are highlighted.

3.3.3 Discussion and gaps in the literature

Various methods for reflection and deliberation are used in the literature. Reflection and deliberation are not uniform interventions and the implementation varies in the different studies. From the existing literature, it is not clear whether there is an ideal structure for the reflection or deliberation exercise. Of the individual-based papers, some focused on encouraging considerations of adaptation while others encouraged participants to reflect on their answer but did not specify what participants should reflect on. It is not clear whether merely asking participants to reflect on their health state values provides enough stimuli for participants to really consider their preferences. The group-based methods also varied in design. Some studies reported that the discussion focused on the implications of participants' valuations while others encourage participants to share their reasoning. Most papers did not report why their specific method of reflection or deliberation was chosen and its content was generally only briefly discussed. Yet, the design of the exercise may be an influential factor. For example, three of the individual-based studies focused on the role of adaptation. An exclusive focus on adaptation may encourage individuals to provide equal or higher valuations for health states, whereas unguided reflection may also reduce health state values. The design of the reflection and deliberation task remains an area for research and is further discussed in Chapter 6.

The varied design of the reflection and deliberation tasks also highlight the potential need for analysis to better understand why participants may value health states differently after reflection and deliberation (Robinson and Bryan, 2013). Two studies (Stein et al., 2006; Robinson and Bryan, 2013) interviewed participants after the task and this may prove useful in better understanding the effect of the reflection and deliberation tasks. Another approach was taken by McTaggart-Cowan et al. (2011), who used a questionnaire, previously designed in a qualitative investigation, to assess why participants' values changed.

Different study designs were used to test the effect of reflection and deliberation. One design is a pre-test post-test design without the use of control groups. In this design, participants are asked to value health states once before and once after the intervention. An alternative design is to compare valuations from a control and an intervention group. The benefit of having a control group is that it controls for other possible causes of change, for example measurement error, but the use of a control group requires extra resources. Both control and no control designs were found in the literature and can be used to test the effect of reflection and deliberation. When eliciting values after the reflection or deliberation exercise participants can be asked to repeat the entire valuation process again or can be provided with their responses and asked to change them if they so desire. Stein et al. (2006) suggest that participants may be hesitant to change their responses if they think that it is more socially desirable to not change their answers. Having the participant repeat the process without knowledge of their original responses may provide a stronger test of whether preferences are complete and pre-existing.

Overall, the evidence on whether reflection and deliberation change health state values is mixed. Using individual-based methods, three studies found changes to the mean health state values but two studies found that it did not change mean health state

values. No clear trend is discernible in the results and further research can provide more evidence on this topic.

From the group-based studies, four found that health state values changed, while one found no changes. The four studies that found changes did not use methods that are recommended by the NICE reference case (EQ-5D valued using TTO (NICE, 2013a)) or the use of conventional methods of aggregation such as mean aggregated values. The study by Krabbe (1996) used an interactive voting system, the study by Akunne (2006) required consensus, the study by McIntosh (2007) used the CLAMES classification system, and the study by Robinson and Bryan (2013) used the PTO. The generalizability of the findings based on those methods to EQ-5D valued using mean of TTO valuations or other methods is uncertain. It was shown that similar methods investigating the effect of adaptation found changes using the PTO (Damschroder et al., 2005b) but not using the TTO and SG (Damschroder et al., 2008). The study that did not find a change was unfortunately underpowered because it had only 15 participants. As Robinson and Bryan (2013) recommend, there is therefore a need for a study that investigates the effect of reflection and deliberation to valuation method such as the TTO and SG. In addition, using the EQ-5D descriptive system would be useful because it is the recommended descriptive system in the NICE reference case (NICE, 2013a).

As argued before, there is the potential of using MCDA as a tool to aid reflection and deliberation. No study was found that used MCDA as a tool for reflection and deliberation. There remains a potential for studying the role of applying MCDA in health state valuation.

Overall, a sizable literature was found on this topic but the findings are mixed and the generalizability to the NICE reference case is unknown. There remains a need for further research on the role of reflection and deliberation in health state valuation.

3.4 Are preferences over health states informed?

The second research question in this review is on studies assessing whether preferences over health states are informed. The PhD student and supervisors were not aware of any paper that assessed whether health state preferences are informed. Searching the Scopus and OVID databases with key words such as "informed preferences" OR "well-informed" OR "patient experience" did not result in any relevant papers. Speculative searching was conducted on Google Scholar, but again no relevant papers were identified. There appears to be a lack of studies in this area, at least in the English language literature accessible using the databases searched.

There is thus scope for a study on developing and implementing methods to determine whether preferences over health states of members of the general public are informed.

3.5 What are the judgements, cognitive processes, and beliefs underlying health state values?

The aim of this section of the review was to survey the existing knowledge about research direction three, namely the examination of the judgements and cognitive processes underlying preferences over health states. In particular, this review focused on the beliefs underlying health state preferences. A list of 12 papers, the valuation method, the qualitative design, method of analysis, sample size, findings, and source is show in Table 3.2. The next sections provide an overview of the literature, discuss the findings, and highlight any gaps in the literature.

3.5.1 Overview of papers

This section does not cover all the findings of the qualitative literature but only the findings relevant to understanding how people value health and which underlying beliefs are important. In most of the studies found on this topic the focus was not necessarily on uncovering beliefs underlying individuals' health state values. Often the qualitative study was conducted with a different aim, which incidentally required some investigation into how individuals value health. For example, Robinson et al. (1997) aimed to explain certain anomalies found in a previous health valuation study, and Osch and Stiggelbout (2007) meant to identify the presence of biases such as loss aversion during health valuation tasks. Identifying the entire thought process and beliefs of individuals while valuing health was not the main aim of those papers. Therefore, the review only highlights the relevant findings from each qualitative paper.

Authors	**Method**	**Design**	**Analysis**	**Sample size**	**Relevant findings**	**Source paper**
Robinson et al. (1997)	VAS, TTO	Semi-structured interviews with think aloud protocol	Thematic	43	• Different thought process for VAS and TTO (e.g. duration in VAS is ignored) • Non-health factors considered ○ Family	Key paper

Baker and Robinson (2004)	SG	Semi-structured interviews with think aloud protocol	Framework and constant comparison	31	• Preferences on the whole well-considered
					• Non-health factors considered
					  ○ Religion
					  ○ Living situation
						Key paper

Devlin et al. (2004)	VAS	Written comments from postal survey	Thematic and constant comparison	≈ 680	• Non-health factors considered
					  ○ Religious and spiritual beliefs
					  ○ Beliefs about euthanasia
					  ○ Access to care and support
					  ○ Happiness
						Cites Robinson et al. (1997)

Damschroder et al. (2005a)	PTO, TTO	Semi-structured interviews with think aloud protocol	Thematic	64	• Equity beliefs important in PTO
					• Adaptation
					• Non-health factors considered
					  ○ Duty to others
					  ○ Enjoyment of life
						Keyword search

van Osch and Stiggelbout (2005)	VAS	Interviews	Thematic	16	• Difficulty explaining choices • Quality of life is considered	Cites Robinson et al. (1997)

Sherman et al. (2006)	RS, TTO, SG	Semi-structured interviews with think aloud protocol	Informal interpretive style	32	• Beliefs about health enhancing one's life
					• Non-health factors considered
					  ○ Current contentment with life and relationships
					  ○ Time-related goals
						Keyword search

McFarlane et al. (2007)	TTO, SG, WTP	Interviews	Informal qualitative analysis	44	• Pivotal life events important for some participants in the TTO	Keyword search

Authors	Method	Design	Analysis	Sample size	Relevant findings	Source paper
Spencer (2000)	SG	Semi-structured interviews with think aloud protocol	Grounded theory	29	• Non-health factors considered o Family and ability to adapt	Cited in Baker and Robinson (2004)

Osch and Stiggelbout (2007)	SG	Think aloud	Thematic	45	• Non-health factors considered
					  o Commitment to children and family
					  o Ability to adapt
					  o Goals
						Cites Baker and Robinson (2004)

Oliver (2007)	Lottery equivalent method	Written in questionnaire	Thematic	25	• Non-health factors considered
					  o Adaptation
					  o Consideration of others
						Cites Baker and Robinson (2004)

van der Pol and Shiell (2007)	VAS, TTO	Semi-structured interviews	Thematic	30	• Extrinsic time goals important in TTO • Coping with illness	Keyword search

Mulhern et al. (2012)	DCE, TTO	Semi-structured interviews with think aloud protocol	Thematic	30	• Wide range of subjective factors considered
					  o Previous health experiences
					  o Effect of health on others
					  o Ability to cope
						Key paper

Table 3.2 Overview of papers included in review on cognitive processes when valuing health

Robinson et al. (1997) investigated the reasons for difference between health state valuations based on the TTO and VAS. For example, they aimed to explain why some health states are valued better than dead on the VAS but are valued worse than dead using the TTO. A total of 43 participants were invited to value health states using both VAS and TTO. While valuing health states participants were asked to think aloud, meaning to verbalise their thoughts. After the valuation participants were asked a set of follow-up questions. The authors concluded that participants tended to consider different factors in the VAS and TTO. In the VAS certain participants tended to ignore duration. The authors note that TTO exercise responses may have been better considered because a wider range of factors was considered. For example, in the TTO more participants tended to mention the effect of health state on their family (e.g. whether they would be a burden).

In a comprehensive qualitative study Baker and Robinson (2004) attempted to investigate whether preferences were 'well-constructed'. A total of 28 participants valued health states using the SG. A think aloud protocol and a semi-structured interview design was used to investigate participants' thought processes. Both Framework and constant comparison method of analysis were used. The authors found that a range of beliefs and judgements were relevant to the participants. Many of these judgements were not about health. The authors note that factors such as previous experience of ill-health, religion, living situation, family situation, anticipated adaptation to health problems, and the need or duty to support others were all prevalent in the thought process of participants valuing health states (Baker and Robinson, 2004).

Devlin et al. (2004) investigated written comments on a VAS valuation postal survey. A total of 1360 respondents returned the questionnaire and about 50% of those questionnaires contained written comments. These comments were analysed using the constant comparison approach. The authors primarily focused on reasons behind

unusable responses, but they also discuss the general approach of participants completing the VAS. The authors find that several non-health factors are considered, primarily religious and spiritual beliefs and beliefs about euthanasia. These factors affect the valuation of the state 'dead' and because 'dead' is an anchor those factors point can have significant effect on the usability of participants' responses. Participants also indicated that they considered certain aspects important but not included in the EQ-5D. These included health aspects such as eyesight, but also non-health state related aspects such as access to care and support and their happiness in life.

Damschroder et al. (2005a) attempted to understand the difference between PTO and TTO valuations by using qualitative methods. A total of 64 participants valued heath states using both the PTO and TTO. A think aloud protocol was used and thematic qualitative analysis was conducted. The authors found that non-health maximising considerations were more common in the PTO than in the TTO tasks and that participants tended to take a more societal perspective in the PTO than in the TTO. The participants mentioned a range of beliefs about the health states. Participants considered their ability to adapt, using methods such as medicine or technology. Participants also considered how the health state would affect non-health factors such as activities, duty to others, and enjoyment of life.

van Osch and Stiggelbout (2005) attempted to build on the study by Devlin et al. (2004) to understand how individuals complete the VAS. Interviews were held with 16 individuals who completed the VAS and then were "asked to elaborate on their approach". The authors identify four approach used by participant: "sort-of", "bisecting the line", "numerical expression", and "small segments". Although these are approaches to completing the VAS it is not necessarily clear how health states are valued within each approach. For example, in the numerical expression approach the participants are said to decide on a numerical value for a health state before drawing a line to the VAS.

But there is little indication on how participants estimate the numerical value. The authors find that some participants found it difficult to explain their choices, perhaps indicating that health state valuations cannot be explained or are a matter of gut instinct. Some participants considered the value of a health state to be related to the 'quality of life' with a health state.

Sherman et al. (2006) conducted interviews with patients with HIV to investigate their understanding of health state valuation tasks. A total of 32 patients valued health states using a rating scale, the TTO, and the SG. The interview method is described as standardised open-ended because participants were asked the same questions but follow-up questions were more open-ended. An "informal interpretative style of analysis" was used. The authors note that the participants' beliefs about adversity enhancing one's life, their current contentment with their current life, and personal relationships influenced their willingness to trade years or risk being dead. Some participants also had time-related goals, for example seeing younger family members grow up.

McFarlane et al. (2007) investigated the use of SG and TTO in 44 dialysis patients. Their qualitative analysis is limited to informal observations as the quantitative aspect of the study was dominant. One important observation was that 11 of 43 patients considered pivotal life events when trading time in the TTO, for example only wanting to live until "a nephew's wedding next summer". This means that the constant proportional trade off assumption of the QALY may be violated for those participants.

Spencer (2000) used a think aloud protocol while valuing health using the SG in order to test the additive assumption of the QALY. In total 29 respondents were interviewed. The analysis method was based on grounded theory. Participants mentioned beliefs about the effect of the health state on their family. Adaptation, by accepting of ill health, was also mentioned as a consideration when valuing health.

Osch and Stiggelbout (2007) investigated the presence of biases such as probability weighting and loss aversion when using the SG. A total of 45 participants were asked to think aloud while valuing health using the SG. Thematic analysis was used to analyse the qualitative data. The authors found that participants considered the effect of ill health on their life goals. Participants also considered commitment to children and family and both physical and emotional adaptation to ill health.

Oliver (2007) investigated the thought processes behind the lottery equivalent method. The lottery equivalent method is a variant of the SG. Twenty-five participants were asked to write reasons for their responses after valuing four health states. Their answers were analysed thematically. The authors found that adaptation and consideration of others were relevant when participants valued health states. One limitation in this study was that the written answers were shorter than answers that could have been obtained using interviews.

van der Pol and Shiell (2007) investigated the presence and effect of extrinsic goals when valuing health using the TTO. Participants were recent mothers because it was hypothesised that they would have strong extrinsic time related goals because of caring for their children. In total, 30 participants valued four health states using VAS and TTO. The health valuation exercise was followed by semi-structured interviews. The data was analysed using thematic analysis. The authors found that indeed some mothers had extrinsic time goals because of their desire to support their children until they reach an age of independence. Participants also considered whether the health state would affect their ability to support the children or make them a burden. In addition, participants considered coping with ill health.

Mulhern et al. (2012) investigated the valuation of health states using the EQ-5D-5L using the TTO and DCE. A total of 30 participants valued health states while thinking aloud. Follow-up questions were asked after the health state valuation exercises.

Thematic analysis was used. The authors reported that a wide range of personal and subjective factors influenced the value of a health state. The factors include previous health experiences, expected ability to cope, and the effect of health on others.

3.5.2 Discussion and gaps in the literature

The literature suggests that a series of non-heath related beliefs and judgements underlie preferences over health states and that the preferences over health states are unlikely to be purely matters of taste. Participants considered factors such as family commitments (Robinson et al., 1997; Spencer, 2000; Sherman et al., 2006; Osch and Stiggelbout, 2007), religion (Robinson et al., 1997; Devlin et al., 2004), pivotal life moments (McFarlane et al., 2007; Osch and Stiggelbout, 2007; van der Pol and Shiell, 2007), the desire or duty to support for others (Damschroder et al., 2005a; Osch and Stiggelbout, 2007; van der Pol and Shiell, 2007; Mulhern et al.), the expectation of adapting and coping with ill health (Spencer, 2000; Oliver, 2007; van der Pol and Shiell, 2007; Mulhern et al.), the desire to avoid being a burden (Robinson et al., 1997; van der Pol and Shiell, 2007), employment opportunities (Damschroder et al., 2005a), and experience of ill health (Mulhern et al., 2012). One consistent finding from the literature is that participants consider their beliefs on a range of non-health factors when valuing health states.

There is less evidence in the literature about how individuals' beliefs affect the valuation process. It is not clear from the existing literature how the non-health factors affect an individual's preferences or what role they play in the process of forming preferences. For example, the family situation may affect an individual's preferences but there is little indication of how the family situation actually affects the individual's preferences. Additionally, any potential interactions between these factors during decision-making are not yet investigated.

There is scope for a study with the main aim of investigating the relationship between the underlying beliefs and individuals' health state preferences. As mentioned before, most studies cited in this review only discussed the underlying beliefs indirectly because they had a different main aim. A study with the main focus of identifying the underlying beliefs and how the beliefs are used in the process of valuing health would be informative. The study could provide a better understanding of the relationship between the health state, the underlying beliefs, and the individual's preferences.

Most papers achieved detailed understanding by using formal qualitative methods, such as a think aloud protocol and semi-structured interviews. Although the analysis method is at times not well described, methods such as thematic analysis, Framework, and grounded theory are often used. Papers that use informal methods, such as informal description of observations made during the task, or that use comment boxes for collecting data provided limited data.

3.6 Conclusion

A literature review was conducted with the aim of reviewing existing knowledge on three topics:

1) the role of reflection and deliberation (including MCDA) in health state valuation
2) the assessment of whether preferences over health states are informed
3) the examination of the judgements, cognitive processes, and beliefs underlying preferences over health states

A literature was found on the effect of reflection and deliberation on health state values. It was argued that the literature on individual-based exercise was inconclusive because some studies did show changes and some did not. Several studies showed that group-based methods did result in changes in health state values. However, most

studies using group methods did not use the most commonly applied health state valuation methods. The ability to generalise findings from those studies to the NICE reference case, for example, is not clear. One study did use conventional choice-based methods but was underpowered. To determine whether preferences can be assumed to be complete a study using reflection and deliberation using conventional choice-based methods would be informative because NICE primarily uses those methods to determine their recommendations on technology adoption (NICE, 2013a).

No literature was available on whether participants are informed about the health states they are valuing. Therefore, a study that develops and implements methods to investigate the topic would contribute to existing knowledge.

A sizable qualitative literature was found investigating the thought processes of members of the public when valuing health. It was established that a set of non-health factors and beliefs are relevant when valuing health and that health state preferences are not likely to be purely matters of taste. There is a gap in the literature describing the relationship between the health state, the underlying beliefs, and the individual's preferences. This was partly because many papers aimed to investigate aspect other than beliefs, for example they meant to investigate aspects of the elicitation task itself. Important aspects of how members of the general public value health remain to be investigated.

A combination of studies addressing the three questions is warranted. The findings can be used to judge whether concerns about assumptions underlying health state valuations are valid. Together, the findings of such studies can expand existing knowledge into the validity of choice-based methods for valuing health and potentially provide directions for improvement of those methods. The next chapter reports on the aim and objectives of this thesis along with the design and methods used to achieve the aim and objectives.

Chapter 4

Aim and study design

This chapter provides an overview of the aim and objectives of this thesis. It also describes and provides the rationale for the methodology, methods, and study design chosen to reach the aim and objectives. This PhD study is made up of three studies. This chapter provides an account of the overall methodological approach. Detailed methods are reported in the chapters reporting on each individual study.

4.1 Aim and objectives

The starting point of this thesis was that methods used to value health are guided by the preference satisfaction theory of welfare. It was then argued that three assumptions of the theory can be challenged, namely the assumptions of complete preferences, informed preferences, and preferences as purely tastes. Alternative viewpoints on these assumptions may be more appropriate for preferences over health states. The aim of this thesis is to test the three assumptions.

The findings of the literature review informed the three objectives and their related research questions. Based on the literature review it was argued that there remain questions about how people value health. There also remains debate about whether preferences can be considered complete because the effect of reflection and deliberation requires more study. There is also a question about how to structure the reflection and deliberation exercise. To better understand the effect of the exercise a qualitative investigation is required. Finally, the literature review in Chapter 3 established that there is a lack of studies investigating whether preferences over health

states are informed. Considering these findings, the objectives and their related research questions are:

I. Objective 1: Describe how members of the general public value health states when using choice-based methods.
    - What are the beliefs or factors that participants consider when valuing health states using choice-based methods?
    - Is there a relationship or interaction between these beliefs or factors?
    - What is the relationship between the beliefs and preferences? How do the beliefs affect preferences over health states?

II. Objective 2: Assess whether there is a need for reflection and deliberation in health state valuation.
    - What is an appropriate reflection and deliberation exercise based on a fuller understanding of how people value health (which should be achieved through objective one)?
    - Are health state values measured before and after a reflection and deliberation exercise statistically significantly different?
    - If so, how important is this difference?
    - Do differences exist at both the aggregate and individual level?
    - What are the reasons for the changes (or lack of changes) following reflection and deliberation?

III. Objective 3: Develop and implement methods to determine whether preferences over health states are informed.
    - How could informed preferences over health states be defined?
    - Is it possible to elicit predictions from members of the general public on their beliefs about health states?

- Do those predictions accurately reflect the experience of patients living in the health states?

4.2 Methodology and study design

This PhD uses a mixed methods approach, where both qualitative and quantitative data are combined to answer a research question (Creswell, 2014, pp.14-15). The combination of both types of data is required because some research questions listed previously need to be investigated quantitatively (are health state values measured before and after a reflection and deliberation exercise statistically significantly different) while other questions need to be investigated qualitatively (what is the reason for the changes after reflection and deliberation). To achieve the objectives and aims of this thesis requires the use of mixed research methods.

The PhD is designed as a sequence of three studies. It can be noted that the presentation order of the objectives in this and the following chapters is different from the order of the assumption in the previous chapters. The order for completing the studies was chosen with the anticipation that objectives two (reflection and deliberation) and three (informed preferences) could benefit from the findings of objective one (how people value health). A better understanding of the beliefs underlying health state valuation could potentially help design a better reflection and deliberation exercise and a better test of whether preferences are informed. Therefore objective one (on how people value health) is studied and presented first. Due to resource constraints the data collection of objective two and three were combined. Part of the data for objective three was collected during the data collection of objective two, therefore the study on reflection and deliberation is presented second and the study on informed preferences is presented last in this thesis.

The three research studies are integrated, meaning that the different components are combined, for example findings of one study are used to design the other studies (O'Cathain et al., 2007; Creswell, 2014, p.230). In this way, the sequential mixed methods design achieves an "integration yield", meaning it achieves more than what would be possible if the three studies were not integrated (O'Cathain et al., 2007). In particular, key findings of study one about the important beliefs underlying health state preferences are used to design the reflection and deliberation exercise of study two and to design methods to test whether preferences are informed. Figure 4.1 presents an overview of the three studies of this study.

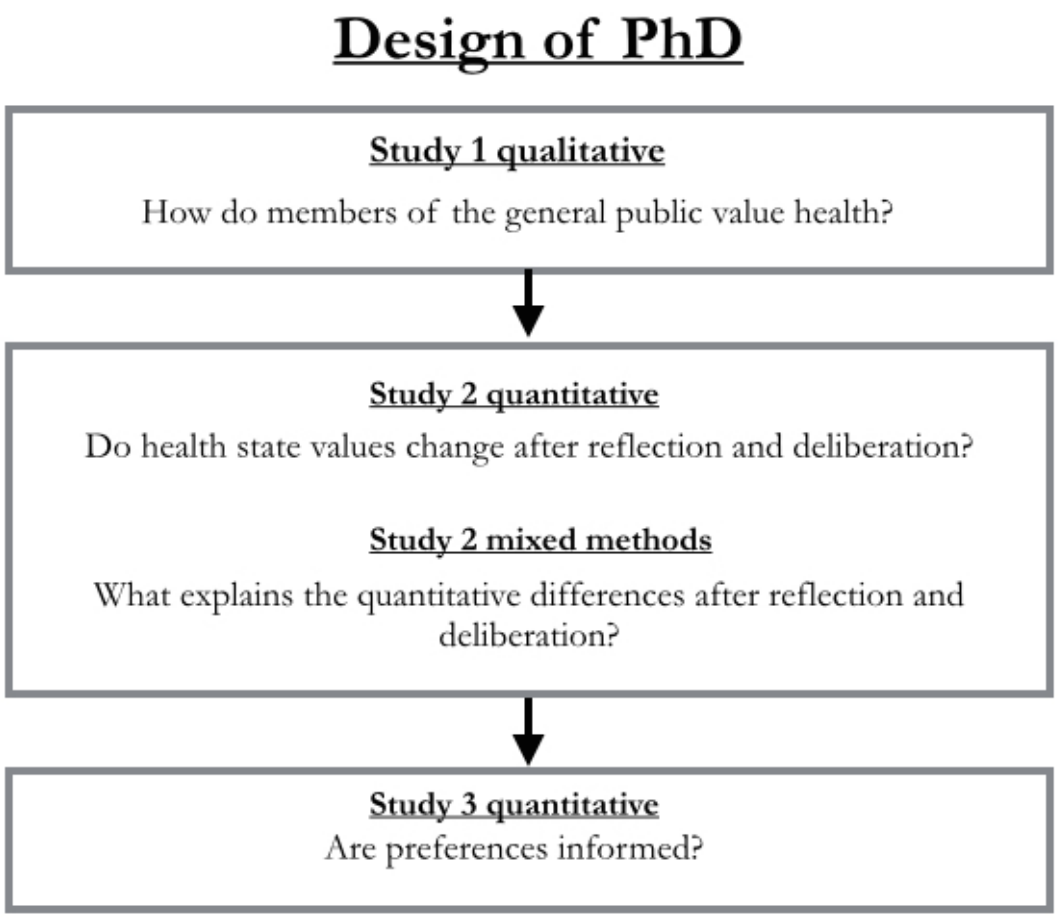


Figure 4.1 Overview of PhD mixed methods design

An overview of the three studies of this study is given in the next sections. Later, detailed descriptions of the methods are reported in the chapters relating to each study.

4.2.1 Study 1: how do members of the general public value health?

The objective of study one is to understand how individuals value heath and to investigate the beliefs that are important when valuing health. This study will aim to answer whether preferences over health states are purely matters of taste or not. Understanding the beliefs underlying preferences can show whether preferences over health states are open to discussion and criticisms or whether preferences over health states, like matters of taste, are not.

Semi-structured individual interviews with a think aloud protocol were used in study one. During the semi-structured interviews participants were asked to value health states using the TTO and DCE with the EQ-5D-5L descriptive system. The use of a think aloud protocol means that participants are asked to tell the interviewer everything they are thinking while performing the health state valuation tasks (Ericsson and Simon, 1993). Follow-up questions were used to probe the participants for more detailed answers. The interviews were audio-recorded and transcribed. The transcripts were analysed using Framework thematic analysis, which is a structured method for the analysis of qualitative data (Ritchie et al., 2003b).

Qualitative methods were chosen because they can help understand how people form preferences over health states (Dolan and Olsen, 2002, p.131). The qualitative tools of individual semi-structured interviews are chosen for the opportunity for in-depth investigation and detailed understanding of personal motivations and decisions (Lewis, 2003, p.58; Ritchie, 2003, p.37). The think aloud protocol was chosen so that the participants' thought processes can be investigated (Ericsson and Simon, 1993). A combination of a think aloud protocol and semi-structured interviews allows for an in-depth investigation into participants' thought processes when valuing health states, allowing the researcher to clarify and explore the reasons behind the values of participants (Ritchie, 2003, p.37).

The findings of this study describe which beliefs and factors are important to individuals when valuing health and explain how those beliefs relate to their preferences over health states. This understanding will be used for two purposes in the rest of this thesis. First, it will be used to develop a reflection and deliberation exercise. The details of this will be explained in Chapter 6. Second, they will be used to develop a method for determining whether individuals' preferences are informed. The details of this will be explained in Chapter 8.

4.2.2 Study 2: is there a need for reflection and deliberation in health state valuation?

The objective of study two is to determine the role of reflection and deliberation in developing preferences over health states and test the assumption of completeness indirectly. Reflection and deliberations provide participants with the opportunity to construct their preferences, more so than in conventional methods that use one-off interviews. Preferences elicited using methods including reflection and deliberation can be compared to those elicited using conventional methods. Investigating the effect of reflection and deliberation on health state values can explain whether conventional choice-based methods that assume completeness and do not offer participants a chance for reflection and deliberation are valid.

Study two consists of two parts, a quantitative and a mixed methods part. In the quantitative part participants valued health states before and after a reflective and deliberative exercise, using a pre-test post-test design (Creswell, 2014, p.160). The reflection and deliberation exercise was designed based on findings from study one and used MCDA. Participants individually completed a MCDA task and engaged in a group discussion about their MCDA responses. The health state values of the participants before and after the reflection and deliberation exercise were compared to see if

reflection and deliberation had prompted any changes. The group discussion was audio recorded and the audio recordings used in the mixed methods part of this study.

The aim of the mixed methods component of this study is to explain the quantitative effects of reflection and deliberation on health state preferences. More accurate implications of the quantitative results can be drawn if the findings are better understood. For example, if preferences do not change after reflection and deliberation it could be that the design of the exercise did not encourage enough reflection and deliberation or because participants did not feel the need to reflect and deliberate.

To implement the mixed methods the quantitative and qualitative components of the mixed methods aspect of this must be integrated (Creswell, 2014, p.230). Integration is achieved with the triangulation protocol (O'Cathain et al., 2010). A triangulation protocol means that the qualitative and quantitative data are first analysed separately (O'Cathain et al., 2010). The researcher develops a set of findings from the quantitative results and one set for the qualitative results. The two findings are then combined and areas of agreement or disagreement between the two findings are explored.

The quantitative component is the responses to the TTO task before and after the group meeting. The qualitative component is the audio recordings of the deliberation section of the group meeting. Qualitative findings were developed from the audio recordings of the group exercise. The audio recordings were transcribed and then analysed using Framework. Based on this analysis, hypotheses were developed that could explain the preference changes of participants during the valuation task. The hypotheses were then tested using quantitative data, which are the TTO valuation collected from the participants. The areas of agreement between the hypothesis and the quantitative findings were explored to see which qualitative findings could explain the quantitative findings. By understanding the reasons for the quantitative findings a conclusion can be drawn about the role of reflection and deliberation.

Some of the data in this study were used in study three. During the MCDA exercise participants provided quantitative predictions of their beliefs about the effect of health states on their life. This data was used in study three.

4.2.3 Study 3: are preferences over health state informed?

In study three the objective is to assess whether preferences over health states of members of the general public are informed. The method used to test whether preferences are informed had to be designed because no existing methods were found in the literature. For someone to be informed, his or her beliefs about health states should be accurate. The beliefs that were tested in this study were those identified in study one. To determine whether beliefs of members of the public about a health state are accurate they can be compared to the experience of patients living in that health state.

Data on beliefs of members of the public about health states were obtained from the MCDA booklets of study two. In study three, aggregate scores from the MCDA task were compared to secondary data on patient experiences. This comparison can show whether members of the general public have accurate beliefs about the experience of patients in certain health states.

4.3 Conclusion

This chapter has provided an overview of the methodology and the study design of this PhD study, which has the aim to test whether assumptions of the preference satisfaction theory of welfare are violated. Three objectives are specified and these were investigated in three studies. The first study describes how members of the general public value health states using qualitative semi-structured interviews with a think-aloud protocol. The second study assesses whether there is a need for reflection and

deliberation in health state valuation or whether preferences can be assumed to be complete. The study investigates the effect of reflection and deliberation on health state values and seeks an explanation of the effect of the reflection and deliberation exercises. The third study develops and implements a method to determine whether preferences are informed by comparing beliefs held by the general public, collected in study two, to the experience of patients in ill health. Together, the findings of this PhD will expand knowledge on the validity of conventional choice-based methods of valuing health.

The next four chapters present detailed description of the methods, results, and discussion of the results for the three studies of this PhD study. The ninth and final chapter of this thesis provides an integrated discussion of the whole PhD study.

Chapter 5

How do members of the general public value health states?

The objective of this study is to describe how members of the general public value health states when using choice-based methods. In particular, this study will investigate the beliefs that participants consider when evaluating health states. Understanding how individuals value health states will answer whether preferences over health states are basic or non-basic, i.e. whether they are purely tastes or not. If preferences over health states are non-basic they may benefit from reflection and deliberation and may be open to rational criticism.

While methods such as SG, TTO, and DCEs are well established in the health economics literature, more needs to be known about how participants complete these tasks. Although previous research has been conducted, in most previous studies the main aim was not to identify the thought processes and beliefs of individuals. There remain gaps in the literature about how individuals' beliefs affect their health state preferences.

In the next section the methods used in this study are described and justified. This includes the interview protocol, the recruitment method, and the qualitative analysis method. In the results section the themes emerging from the analysis are reported and supporting quotes for all themes are provided. The themes are then used to develop an explanatory framework. Finally, the findings are discussed and the implications of those findings for health state valuation are considered.

5.1 Methods

Qualitative semi-structured interviews using a think-aloud protocol were conducted. The participants completed both TTO and DCE tasks, and the EQ-5D-5L descriptive system was used to describe the health states. The interviews were transcribed and analysed. The School of Health and Related Research ethics committee approved this study (see Appendix 2). The details of the methods are described in the following sections.

5.1.1 Interview protocol

The study used a semi-structured interview method along with a think-aloud protocol, which is implemented whilst participants completed the valuation task. The semi-structured interview format was chosen because it provides an opportunity for in-depth investigation and detailed understanding of personal motivations and decisions (Lewis, 2003, p.58; Ritchie, 2003, p.37) while the semi-structured aspect of the interview also allows for exploration of a relatively under researched area (Arthur and Nazroo, 2003, p.111).

In order to understand participants' thought processes a think aloud protocol was used (Ericsson and Simon, 1993). A think-aloud protocol means that participants were asked to say out loud everything they were thinking while completing each task. Both retrospective and concurrent think aloud protocols are available. A concurrent think-aloud protocol was used, which means that data is obtained while respondents are completing a task. This avoids participants having to recall their thoughts after a task has been completed, which may result in them adding information generated after the task had already been completed (Ericsson and Simon, 1993, pp.19-30). All think aloud protocol guidelines were developed by Ericsson and Simon (1993). If participants became quiet during the interview, they were reminded to "please keep talking". This

prompt is recommended because it is less directive than other possible prompts (Ericsson and Simon, 1993, p.256). If participants had not verbalised their choice or had done so only minimally, they were asked to "repeat everything you can remember about your thoughts".

The semi-structured interview consisted of several stages. Each interview started with a practice task of choosing between two cars, in order for participants to become familiar with thinking aloud (see Appendix 3). After the practice task participants valued health states using both the DCE and the TTO method. Follow-up questions were asked after the valuation task. Questions were asked about what factors were important in participants' choices and participants were given a chance to comment on the valuation tasks. Occasionally questions were asked during the think-aloud protocol (before the end of the entire valuation task) but were generally avoided because they could potentially influence the participants' thought processes for the health states which were yet to be valued. These questions were based on participants' comments and were meant to clarify their comments. They were only asked when the interviewer thought waiting till the end of the interview would mean that the interviewee would not be able to answer the question. At the end of the interview, a questionnaire was used to collect standard background information, such as age, gender, and marital status. The background questionnaire is shown in Appendix 4. The interview topic guide is displayed in Table 5.1.

All participants signed consent forms, indicating agreement to take part in the interview and consenting to being recorded. Participants were reminded that they could leave the study at any stage without any penalties. The audio recordings were stored on a password protected encrypted drive and the transcripts were anonymised and stored in a locked filing cabinet on university premises.

Stage of interview	**Interviewer guidelines**
Introduction	Provide and discuss information sheet and consent form Ask permission to record
EQ-5D form	This form asks questions about your health as it is today
Explain think aloud	In this experiment we are interested in what you think about when you find answers to some questions I am going to ask you to answer. In order to do this, I am going to ask you to think aloud as you work on the problems you are given. What I mean by "think aloud" is that I want you to tell me everything you are thinking from the time you first see the question until you have given your final answer to the question. I don't want you to plan out what you say or try to explain to me what you are saying. Just act as if you are alone in the room speaking to yourself. It is most important that you keep talking. If you are silent for any long period of time, I will ask you to talk. Do you understand what I want you to do?
Practice task	Ask participants to choose between car A and car B
DCE Explanation	I want you to imagine that you living in either of these states. You would be in these states even if you were receiving treatment. I would like you to choose between health state A and B. Remember to think aloud.
DCE x 8	Here is a new choice, please remember think aloud
If not talking for 20 seconds	Please keep talking
If not talked a lot after a choice	Please report everything you can remember about your thoughts during the last choice
Possible follow up questions	Could you tell me how X played a part in your choice? Why did you think this would be affected in such a way What made you think X would be this way?
Explain TTO	Now this is a slightly different task. I'm going to ask you to make a choice between the state at the top, which is life A and the state at the bottom, which is life B. The choice is between being in life A for a certain length of time, and being in life B for 10 years. From now on, I want you to imagine that you yourself are in these states, even if receiving treatment for the health state, and they would last for up to 10 years without any change and then you would die.
TTO x 3	Here is a new health state, please remember to think aloud
Follow up end interview	What were your main considerations when giving your answers? Was there anything you didn't mention that you think is important? If you had more time could you please tell me what other information would you like to know or consider before you answered the questions. Is there any reason you did not mention (list factors) during the choice tasks? Would considering this have changed your choice?
Background	Background questionnaire to fill out.
Conclude	Thank and debrief

Table 5.1 Topic guide for interview

EQ-5D-5L and chosen health states

Health states used during the interviews were described using the EQ-5D-5L (Herdman et al., 2011). The EQ-5D-3L descriptive system (Rabin and de Charro, 2001) is a frequently used questionnaire and is included in the NICE reference case (NICE, 2013a). The EQ-5D-5L is a newer version of the EQ-5D-3L and is meant to be more sensitive and responsive because of the increased number of levels (Herdman et al., 2011). The EQ-5D-5L is mentioned in the NICE recent guidelines (NICE, 2013a) and may be adopted in the future. The EQ-5D-5L was chosen so that this research is relevant when the 5L version is adopted and because there are not many qualitative studies on the 5L version. The EQ-5D-5L contains five domains: mobility, self-care, usual activities, pain or discomfort, and anxiety or depression. Each domain contains five levels, describing a severity of problems. For example the mobility domain ranges from "no problems in walking about" to "unable to walk about". The EQ-5D-5L is included in Appendix 5.

The health states were selected in order to present participants with problems in each domain and with a range of severities. The health states included those valued worse than being dead according to the preliminary value set (van Hout et al., 2012). States worse than being dead were included in case the participants' thought processes were different for milder states and because worse than dead states in the TTO use a different elicitation procedure (see page 74).

Each participant completed eight DCE and four TTO tasks. The number of states was chosen to allow the interview to be completed in one hour. More DCE states were chosen than TTO states because the DCE task takes less time per state. The order of the DCE and TTO tasks and the order of the states within each task were randomised. This was done because there is the possibility that participants' thought processes may be influenced by the order of the health states. For a list of states used see Table 5.2 and

Table 5.3. The numbers represent each health state and are used in the text throughout this thesis. Each health profile of the EQ-5D can be described by a five digit number, with each digit representing the level of each domain. The numbers represent the levels of each domain from best to worst (i.e. 1 is no problems and 5 worst outcome level for the EQ-5D-5L) and the order of the digits is the order of the domains in the questionnaire. For example, in the EQ-5D-5L the number 13321 represents the health state of no problems on mobility and anxiety or depression, moderate problems on self-care and usual activities, and slight problems on pain or discomfort. The full description of the states is given in Appendix 6.

DCE choice task	**DCE State A**		**DCE State B**	
	Year	State	Year	State
1	10	33243	10	33234
2	10	21221	10	12131
3	10	13323	10	31332
4	8	43312	10	33411
5	5	34454	5	43544
6	10	23211	10	12311
7	5	33341	5	53321
8	8	22432	10	22233

Table 5.2 Health states used for the DCE during interview

TTO task	**TTO state**
1	13321
2	13443
3	54435
4	31212

Table 5.3 Health states used for the TTO during interview

Both DCE and TTO were used because both techniques were explored in the valuation of the EQ-5D-5L (Devlin and Krabbe, 2013). The techniques are described below.

The discrete choice experiment task

In the DCE task participants were asked to choose between two health state profiles. Participants were forced to choose and no opt-out was available. This is the method investigated for the EQ-5D-5L valuation and often used in the literature to develop

health state values (Stolk et al., 2010; Oppe et al., 2014). An attribute for duration was included in all profiles, but was only varied in some states. This was done in order to avoid cognitive overload. The DCE task was implemented as a self-complete booklet using pen and paper because it was less resource intensive to produce than a computer-based DCE. For the format of the DCE see Appendix 7.

The time trade off task

The TTO task used in this interview is the 'composite-TTO' (Devlin and Krabbe, 2013). The conventional TTO approach was used to elicit the values for states better than dead, while lead-time TTO was used for health state values worse than dead (Oppe et al., 2014). This was done because the EQ-5D-5L valuation may use composite-TTO (Oppe et al., 2014). In the conventional TTO, the time frame was 10 years. In the lead-time approach, life A was composed of 15 years of full health followed by being dead and life B was composed of 15 years of full health followed by 10 years in the health state to be valued. The task was implemented using two TTO boards (see Appendix 8).

The procedure used for arriving at indifference for states better than dead was based on the Measurement and Valuation of Health (MVH) protocol (Gudex, 1994). The participant was first asked to choose between 10 years in full health and 10 years in the health state to be valued. Then the choice was between zero years in full health (immediate death) or the health state. If the participant considered the state better than dead the marker was moved to 5 years. After this the marker was moved by increments of one year either to the left or right until a point of indifference was reached. For states considered worse than dead, the participants compared life A and life B. If the participant selected life A, the number of years in full health was reduced to zero. If the participant then selected life B, the number of the number of years in full health for life A was moved to eight. After this the years were moved by increments of one year until indifference was reached.

5.1.2 Pilot

Two pilot interviews were conducted. Based on the pilot interviews slight changes were made, for example the warm-up task was changed. The original warm-up task suggested by Ericsson and Simon (1993, p.376) was an anagram where participants were instructed to find an English word from a set of scrambled letters. In the pilot interviews the anagram task put undue stress on the participants who were not able to solve the task. In addition, the task was not related to the DCE or TTO tasks. Instead, a warm-up task of choosing between two automobiles in the DCE format was selected. The overall structure of the interview was deemed appropriate and no other major changes were made after the pilot.

5.1.3 Sample size and recruitment

The total target recruitment number was 20 participants. Sample size in qualitative research is generally smaller than quantitative work because there is a faster diminishing return, a lack of need for statements about incidence or prevalence, and rich detailed data for each participant (Spencer et al., 2003, p.83). The sample size was chosen after reviewing similar research projects, which ranged from 25 (Oliver, 2007) to 45 (Osch and Stiggelbout, 2007), and considering practical resource constraints.

Participants were recruited from three sources: a list of Sheffield community organisation groups, a list of student and staff from the University of Sheffield, and using the snowball method. The 'Sheffield Help Yourself' website (Sheffield Community Information Service, 2015) was used to contact Sheffield community organisations. The website listed a contact person for different community and social groups. The contact persons of several groups, chosen at random, were contacted by email with a request to forward a recruitment email to all group members. The student and staff from the University of Sheffield were contacted by email using a University

managed emailing list. Academic staff and ScHARR staff were excluded because they are likely to be less representative of the population and more likely to have knowledge of the health valuation techniques. The snowball method means that participants in the study assisted in identifying other potential participants by asking them to share details of the study with others they thought may be interested (Ritchie et al., 2003a, p.94). In this case either interviewees identified other participants or recipients of the recruiting email forwarded the invitation email to acquaintances.

Participants were purposively sampled over age and gender. There is some evidence that age (Dolan and Roberts, 2002; Wittenberg et al., 2006; Essink-Bot et al., 2007) or subjective life expectancy (van Nooten et al., 2009), and gender (Dolan and Roberts, 2002) are related to people's health state valuation. This means it is possible that age and gender may influence people's thought processes and beliefs considered in valuing health. Although other characteristics are associated with people's preferences, such as current health status (Dolan, 1996) and marital status (Dolan and Roberts, 2002), this study could not include a sufficient number of participants to further split the sample. Each participant was reimbursed with a £10 gift voucher.

5.1.4 Analysis method

Framework (Ritchie and Spencer, 1994; Spencer et al., 2003) was used to analyse the transcripts. Framework was chosen for several reasons. First, it is appropriate as the aim of the research is to develop an explanatory account rather than more complex forms of qualitative analysis, such as those focusing on life history, interaction between participants, or documentary analysis, which are better analysed using methods such as narrative, conversation, and discourse analysis (Gale et al., 2013). Framework was developed for applied policy questions and provides a structured analysis process and produces highly structured output (Gale et al., 2013). Framework helps the researcher

by providing a systematic method of categorising, organising, and analysing qualitative data, and provides outsiders with a clear outline of what analysis steps have been followed.

Framework analysis involves several stages: familiarisation, identifying a thematic framework, indexing, charting and mapping, and interpretation (Ritchie and Spencer, 1994, p.178; Ritchie et al., 2003b). These steps were followed in this study. First, each interview was audio recorded and transcribed verbatim by the PhD student. Then during the familiarisation stage, the PhD student familiarised himself with the data by listening to the recorded tapes and reading the transcripts. This was done to develop a list of key ideas and recurrent themes occurring in the data. The key ideas and themes were used as a basis for conducting a detailed thematic analysis.[10] A thematic framework was developed based on the list of recurrent themes. Themes were then sorted and grouped according to their content. After this process a list of main themes along with their sub-themes was created. All the transcripts were then indexed based on that thematic framework, meaning that sections of the transcripts that refer to a theme in the thematic framework were identified and marked. This process was conducted several times and in an iterative fashion. All transcripts were read at least three times. To chart and map the data, a matrix of all themes and participants was created, listing the themes by columns and the participants in the rows. Thus each cell represents space for comments from one participant on one sub-theme. For each cell in this matrix the relevant comments, if a comment was made, were entered. Using these charts, a descriptive and exploratory analysis can be conducted (Ritchie and Spencer, 1994, p.178).

---

10. Researchers can use pre-existing ideas or themes when developing the thematic framework but this was not done in this study. The goal was to approach the data without any pre-conceived theoretical framework so that primarily the data would determine the findings.

During the interpretation stage a descriptive and an explanatory account are developed. Descriptive accounts focus on "identifying key dimensions" or "mapping the range and diversity of each phenomenon" (Spencer et al., 2003, p.214). In the explanatory account the researcher attempts to find "patterns of association within the data and then attempt to account for why those patterns occur" (Spencer et al., 2003, p.215). The aim of the explanatory account is to link the themes together to explain some phenomenon, for example in this PhD study the aim is to explain how people formed preferences over health states. Linkages can be mentioned explicitly by the participants, but if they are not explicitly stated the researcher must look for the linkages by examining whether certain themes are closely associated in the data (Ritchie et al., 2003b, p.253). In addition, 'common sense' can dictate how these themes hang together "and may provide a possible underlying logic to what people have said" (Ritchie et al., 2003b, p.253).

The explanatory stage of the analysis is further removed from the data than the descriptive stage and therefore requires more interpretation by the researcher. Furthermore, an implicit explanatory account requires more interpretation by the researcher than an explicit explanatory accounts because the links between the themes are not identified by the participants but by the researcher. To verify whether a chosen implicit explanatory account is justified by the data further analysis can be conducted. Miles and Huberman (1994, p.263) suggest three techniques to verify explanatory accounts in qualitative research. These three techniques are:

- **Look for outliers.** Miles and Huberman (1994) recommend to look for cases, or settings, that are outliers to the conclusion that is drawn. This can be done by asking, "are there exceptions to this finding?"
- **Look for negative evidence.** This technique is "a more extreme version of looking for outliers" since disconfirmation is sought (Miles and Huberman,

1994, p.271). The researcher asks, "do any data oppose this conclusion or are inconsistent with this conclusion?"

- **Find and test rival explanations.** Here, the analyst develops alternative explanations and assesses those. This involves asking what the next best explanation could be or what could disprove the explanatory account.

5.2 Results

In total 21 semi-structured interviews were conducted. Two participants were recruited from a list of Sheffield community organisation groups. Fourteen participants were recruited from a list of student and staff from the University of Sheffield. Five participants were recruited using the snowball method. The description of participants' background characteristics is shown below in Table 5.4.

	Sample	UK population[1]
Number of participants	21	-
Female (%)	11 (52%)	50.8%
Average Age	42	39
Age range	19 to 65	-
Age bracket 18 to 29	6 (29%)	20%
Age bracket 30 to 49	6 (29%)	36%
Age bracket 50 to 59	5 (24%)	15%
Age bracket 60 above	4 (14%)	29%
First degree or higher (%)	13 (62%)	27%[2]
Student (%)	6 (29%)	8.8%
Employed (%)	14 (67%)	59%
Retired (%)	1 (5%)	13%
Previous illness (Personal, Family, or Friends) (%)	16 (76%)	-
Average EQ-5D score (%)[3]	0.88	-
Range EQ-5D score (%)	0.479 to 1	-
Single (%)	12 (57%)	-
Married (%)	8 (38%)	-
Divorced (%)	1 (5%)	-
Children (%)	6 (29%)	-
Religious (%)	9 (43%)	-

1: (Office for National Statistics, 2011), 2: Includes everyone 16 and above, 3: (van Hout et al., 2012).

Table 5.4 Sample characteristics

Themes identified from transcripts

The Framework analysis identified four main themes to describe how participants value health states. These themes are:

- **The interpretation and concretisation of a health state** – Any comments relating to how a participant interpreted the health state, including how they converted abstract health states into concrete health problems

- **Conversion factors** – Defined as any comments relating to the how a participant's personal and social situation affects their health state preferences
- **Non-health consequences of health states and the weighing of the consequences** – Defined as any comments about the effects of a health state on the non-health aspects of a person's life and about how important the effects are to the participant
- **Task related themes** – Defined as any comments specifically about the TTO and DCE tasks

Each of these themes is composed of several related sub-themes. The themes and their sub-themes are described below. In the final two sections, the themes are linked together to form an explanatory account of health state valuation. This explanatory account is then evaluated.

5.2.1 Theme 1: the interpretation and concretisation of a health state

Participants spent time interpreting the EQ-5D health states. The EQ-5D contains an abstract description of a state, such as "moderate problems in walking about". The participants generally converted the abstract description into a more concrete feeling or image:

> *I see somebody as [having] severe [problems walking about] who is probably missing a leg. (Participant 15)*

The next sections describe the various findings about this theme, first by providing examples of how different health domains were interpreted.

Examples of participants interpreting different domains

Mobility. Various examples of mobility were provided. For example in this quote 'slight' mobility problems was described as slowing down the process of walking:

When it says slight, I don't imagine that you wouldn't be able to do it. I just imagine that it, I don't know, takes longer (Participant 17)

In the following example 'severe' mobility problems was described by the need to use a wheelchair:

If it's severe I, I would kind of think of that as perhaps you can only walk very short distances, or you maybe need to use a wheel chair (Participant 4)

Self-care. Slight and moderate problems are explained as concrete problems during the washing or dressing routine in the quotes below:

I mean, not sure, slight problems I can quantify that as say well, like trouble getting the toothpaste cap off (Participant 3)

Moderate so perhaps it just means for example you can't easily put your own shoes on (Participant 1)

Usual activities. Various definitions were used for this domain. Some people focused on a broad range of activities while others focused on very specific activities:

I would have some problems doing usual activities, so it might be doing something like doing the gardening, I wouldn't be doing it for so long, wouldn't be doing a lot of digging (Participant 7)

Severe problems doing your own usual activities, would to me mean that you're not even getting out in society, you are house bound, you know you can't get to work, you can't get to, you know, take your children to school, you can't go to, uhm, the shops, you can't do any kind of hobbies (Participant 11)

Pain or discomfort. Pain was often compared to previous episodes of pain, such as injuries, surgical procedures, and childbirth:

I've never really been in severe pain. Apart from childbirth (Participant 7)

Anxiety or Depression. Concrete descriptions of anxiety or depression were rare. For some, anxiety or depression was compared to moments of stress:

Slight anxious or depression is, that's when you got way too much work going on or you know suddenly, yeah too much pressure (Participant 20)

The previous examples illustrate how participants interpreted the domains and the labels of the EQ-5D-5L attribute levels. The interpretations tended to be more concrete than the EQ-5D labels, as in the example of severe pain being interpreted as the type of pain that occurs at childbirth.

Imagination and experience

In interpreting states, participants used their imagination and their experience with health problems. These quotes illustrate the various sources of experience people used in imagining the health state:

Well, I mean I think uhm, I mean my mother is [age] and she does have problems washing and dressing herself. (Participant 7)

I had an operation at one point, and when I came out of the anaesthesia I was in extreme pain. (Participant 4)

Participants referred to several sources of experience, such as personal experience; experience of family, friends or acquaintances; work experience; and knowledge of the experience from celebrities, the news, or documentaries. Although many participants could rely on their experiences for many health problems, occasionally participants had to imagine what a health state would be like:

I mean I haven't suffered most of these problems myself nor fortunately have I had relatives or friends who had them really. Uhm, so really it's just, I'm just I'm using my imagination. (Participant 10)

Both imagination and experience were important in participants forming their concrete interpretation of the health state.

Adding information to the health state

Some participants seemed to read more into the health state than was written. This involved trying to guess the disease that caused the health state or finding out what is 'implied' or 'suggested' given the information in the health state:

> *I would suggest that, that would imply that you've got, if you've got no problems in walking about, you've got moderate problems washing and dressing yourself, that would imply to me that this this health state involves some kind of loss of upper body mobility uhm, what that might be I don't know. (Participant 9)*

> *I think walking about is possibly more important, in that, that suggest to me something about overall physical kind of capacity. (Participant 5)*

These quotes illustrate participants adding information to the health state when interpreting the health state. Their statements illustrate how they sometimes focus on what is 'implied' or 'suggested' given the information in the health state.

Lack of experience

Participants occasionally expressed doubt about their ability to accurately imagine what a health state is like because of their lack of experience with the health problems in that state. This meant that it was difficult for them to know what the health state would concretely involve:

> *Perhaps you would get a better handle on this if you spoke to people who did have some of these problems, cause I mean I'm imagining what it's like, and I've got some slight insight what it's like to have pain or slight insight into what it's like to not to, but I have not really because I'm happy today, I've been happy all my life. I mean I've been fit all my life pretty much so. It's very, I I I find it very difficult to say just how much these things, would cause me to actually trade years of life of. (Participants 10)*

> *It's really really hard this one, because I think, uhm, I don't see myself as an anxious or depressed person and I've never had any big bouts of depression so it's really hard for me to empathise. (Participants 11)*

These quotes illustrate that a lack of experience makes it difficult for participants to concretely think about health issues. But participants also spontaneously expressed

doubt about the lack of experience of others. Four participants mentioned that others would think differently about anxiety or depression:

> *I think mental health is like one of those things that unless you know someone, or you've experienced it, that you'd probably have no concept of. So may be physical problems might come to the fore, for some people" (Participant 4)*

Not believing states

Occasionally certain states were not believable to some participants. This often involved perceived inconsistencies between the domains of mobility, self-care, or usual activities but also between pain or discomfort and anxiety or depression domains combined with usual activities:

> *Those are likely to go together aren't they? Can you have moderate problems walking around and no problems with washing and dressing, I'm not sure about that. (Participants 10)*

> *I would, I would contend that this is a bit, this is a bit confused, cause you say that I've got more problems uhm uhh with my day to day activities in life B, where even though I got no problems in walking about. (Participant 16)*

> *In this one you say you have slight problems doing your UA, but you're extremely anxious or depressed. Well I've already said that causes more problems for doing activities than physical problems, so I don't know, you know, I'm just querying the actual states you've created" (Participants 6)*

> *If you have moderate pain or discomfort, how would you have no problems doing your usual activities? (Participant 4)*

Rewriting health states

As a result of not believing health states, participants sometimes 're-wrote' the health state. This meant that they explicitly changed one of the levels for a given domain into a different level that they felt was more likely:

> *So even though it says that I'm not anxious or depressed, the outcomes of these two problems would actually make me depressed. (Participants 13)*

I would contend that possibly not being able to do my day to day activities would make me depressed, hehe, in fact I know it does so uhm I would probably go for life...B (Participants 16)

Another way of re-writing health states was by adding treatment. Despite instructions to imagine living in the health state of the specified duration, some participants added treatment:

For physical pain, you can take medication and again become better. (Participant 13)

Severe pain and discomfort, there are medicines available to deal with that, and uhh obviously there are things like pain clinics, and referrals from your GP to go and deal with that. Uhh. Moderately anxious or depressed I would say definitely look at uhh some perhaps stronger or different medication to cope, to deal to deal with that. (Participant 19)

Participants tended to add treatment to certain domains more than others. Thirteen people mentioned treating 'pain or discomfort' or 'anxiety or depression'. Only two participants mentioned treating mobility problems.

This section illustrated the complex way in which participants interpreted the EQ-5D-5L health state descriptions. Using their imagination and experience the participants concretised the health states. The process of interpreting was not straightforward because participants expressed doubt about their knowledge, added information to the health state, did not believe some health states, and even re-wrote some health states.

5.2.2 Theme 2: conversion factors

Conversion factors are personal and social factors that affect how participants value health states. The term 'conversion factor' was chosen as these factors represent an intermediate step between the health state and the consequences of the health state.[11,12]

---

11. The term 'consequences of the health state' is discussed in the next section
12. The term 'conversion factors' was borrowed from the Capability Approach. Sen uses the term to convey factors that turn goods into their end means (Sen, 1987, pp.6-11). For example, the ability to bike converts a bike into a mode of transportation.

Conversion factors were grouped under four categories: (1) personal interests and circumstances, (2) other people's reaction to ill health, (3) the ability to adapt or change expectations of life, and (4) available support.

Personal interests and circumstances

In some cases the personal interests and values of a participant determined the effects of a health state. The same health problems with self-care or mobility can cause different consequences depending on a participant's personal interests and values:

> *Well for me, personally, I find, I'm a runner for a start, so if I've got walking problems I couldn't run...which would upsetting. (Participants 16)*

> *Again uhm, I think I'm not someone who has a shower every day, I think there's a bit of a fetish about cleanliness. (Participants 6)*

The same mobility problems can have different values in different circumstances, such as if one lives in the countryside rather than in a city:

> *I think some of these things, which you find when you ask other people it's very subjective upon how that would impact my personal life on a day to day basis. Uhhm. And that would constantly change, when I go back home to see my family. Because mobility becomes less of an issue because you're stuck in the countryside and nobody sees nobody for a couple weeks. (Participants 15)*

These quotes illustrate how participants reflected on the health state in combination with their own personal interest and circumstances. The personal interest and circumstances help determine the effect of the health problems on someone's life.

Other people's reaction to ill health

Participants considered the social environment and how people in wider society view different health problems:

> *Having emotional problems seems to be uhhmm, a negative thing in the world. With emotional problems, people usually think you are less able. (Participant 1)*

I suppose some people are embarrassed about having to use wheel chair or something like that, and uh I don't think these days, you know people single people out as much, if you've got physical disablements, you know, people don't single you out as much. (Participant 8)

Participants considered their social environment and the outcome of a health problem was dependent on how other people in wider society may react to different health problems.

Adaptation or change of expectations

Some participants stated that they expected to adapt to a health state or change their expectations of life. The following quote illustrates how participants indicate they could adapt by changing their usual activities:

If I did have uh problems doing my usual activities, then most of my usual activities are quite sedentary so I may then choose perhaps something else to do. Because they tend to be craft based so I'm thinking, well if I couldn't perhaps do intricate craft work anymore, maybe I could do patch working or you know, I can move on to something that was less intense than doing tiny cross stitching things like that (Participant 14)

By moving on to other activities, the participant lessened the effect of the health state. Some participants mentioned adapting their routines and making accommodations in their life to adapt to a health state:

I think, once you have, know you're way around certain things and you learn which where the uhm, none walking access things are. So probably lifts and which buildings have wheel chair assisted doors uhm, then you make accommodations in your own life and you have can a fairly sort of, what becomes a normal life. (Participant 11)

Some participants acknowledge the possibility of changing their expectations of life, for example by accepting their problems:

Acceptance, you gotta accept that you got problems. (Participant 12)

You have to think that you know...although you might not be used to, you will get used to it you know in time. You have to think the carers coming in to you, to see you, are used to it. You know what I mean, I suppose you would get used to it. (Participant 8)

However, some participant's struggled with deciding whether they would be able to adapt to ill health:

> *(...) you're rendered tetraplegia and you can't, you just lie there paralysed and you can't do anything you know. And yet, you hear and see people who apparently find a fulfilling life under those circumstances. I can't imagine that I could, but people seem to. Whether that's just because like so many of us they're just so frightened about what death is, if it's anything at all, that any any any life, however poor quality, is better than none, that may be it. But maybe they do actually find some fulfilment in what appears to be a very wretched life. (Participant 10)*

Participants considered adapting to a health state, perhaps by changing their activities, but also by changing their expectations, perhaps by simply accepting their problems.

Available Support

The last conversion factor was the support available to an individual. Participants considered many sources of support: family and friends, charities, government support, caretakers, the physical and social environment, available technologies, and treatment possibilities.

Some participants considered whether, and how, their family and friends could support them with their ill health:

> *Well I think that if you're kind of dealing with hmm health problems like this, and if you're kind of, uhm uhm I think you have to kind of make a personal decision based on uhm who you have. If you've got a good support network in terms of family and friends, and whilst if you've got quite a small people and also those people that you do kind of have, the family, you can't rely on them for things anyway. So that, you'd kind of have to, kind of balance all that into the equation." (Participant 4)*

Help from charities was also considered:

> *I'm thinking...of drawing on...what probably charities linked to whatever kind of medical condition I had could provide. (Participant 11)*

Religion was occasionally mentioned:

I think maybe because that relates to my my uhmm my religion, I guess. Because I am a [religion] so I guess...even if you are anxious or depressed, I can (...) I just pray about it. (Participant 2)

Government and NHS services were mentioned:

They've got a home help who comes once a week but that's it, that's the only outside source of help at the moment the NHS is providing to them (Participant 11)

Many considered care takers helping them, as well as how they would feel about this type of care:

I don't like, cause of what I hear about people who have carers coming in, I don't like the thought of that. Because I got a friend whose son did that job for a while and you're told, you know, you've got 15 minutes here, that's all. You can have and you know it takes longer than that, so you can choose between either getting hopelessly behind with your work or just doing a bad job. And I just I just uhm, yeah I don't like the sound of that really. (Participant 5)

For some participants the physical and social environment was important in determining the impact of different health problems. This involved considering factors such as transportation, disabled access, the law and the work environment:

The society where I live at the moment, there does seem to be a lot of I mean just around the university there's a lot of accommodation for people with disability and there's a lot of focus, and there seems to be quite a lot of help and assistance. So I'm probably quite lucky to be in that environment, and I almost think that if I had, moderate problems walking about, there would be things in my workplace put in place to accommodate those things. (Participant 11)

I think it's uhhm also regulation uhhm and and the law really that helps disabled people or people with mobility difficulties. (Participant 19)

Technology was also considered as a support tool. The use of technology and treatment both refer to medical interventions to manage the health problems. The difference is that technology does not alter the health state while treatment does.

Considering treatment is therefore not correct in valuing a health state, while considering technology is[13]:

> *I'm just thinking, say it had been a problem with walking, then you'd look at mobility scooters or you'd look at wheel chairs or whatever, so there is sort of technical aids. (Participant 6)*

> *(...) if you end up in a, well worse you can't, you can't even be in a wheel chair, you end up bed bound or something. I mean particularly nowadays with with computer technology, you can still communicate with people, you can do all those things, you can can be happy. (Participant 6)*

As a final aspect of conversion factors, it was noted that the ability to consider conversion factors depended on the knowledge that participant had:

> *You look at what's in the news...and then it's quite publicised that things like depression, or any kind of mental illness, there's not a lot of help out there for. (Participant 11)*

When evaluating health states participants considered a set of conversion factors, which determined the way in which ill health affected them personally. A range of conversion factors were considered, such as an individual's personal circumstances, other people's reaction to ill health, an individual's ability to adapt or change their expectations, and the available support. Participants had to consider which conversion factors were relevant for them and how different health problems would interact with different conversion factors. The conversion factors helped determine what the consequences of a health state would be for each individual.

5.2.3 Theme 3: non-health consequences of health states and the weighing of the consequences

Participants considered the consequences of a health state for their life. The consequences of a health state are the non-health effects that are caused by ill health.

---

13. Example quotes for the adding treatment sub-theme were given in section 'Rewriting health states' on page 86.

Participants mentioned many non-health consequences of health problems. Health problems could have a range of consequences, for example on an individual's independence, dignity, ability to support others, level of burden imposed on others, income, and purpose in life. The next section describes the various sub-themes related to this theme, first by providing examples of consequences.

Examples of consequences

Although choices between health states often seemed to be between health domains, it was actually based on a deeper consideration of non-health consequences. For example, the first quote illustrates that although the choice between life A and life B was initially explained as a choice between mobility and self-care, the choice between mobility and self-care was in turn based on a consideration of independence:

> *I would probably choose life B.... having the ability to walk about, uhh but also keeping yourself clean and tidy and washing and dressing yourself uhhm. Having severe problems with that would would outweigh I think...uhhm well I think uhm from a personal hygiene point of view uhm if you're able to uhm keep yourself clean and dress yourself, uhm I think that outweighs having severe problems with walking about uhhm...there are aids that can be used, wheel chairs uhm, frames, sticks etc, uhhhm, and you would need assistance if you had severe problems washing and dressing, so it's more of a more of looking at it from an independence point of view, from having your own independence, that's how I look at it. (Participant 19)*

The quote illustrates how the preference between health domains is based on how ill health affects independence, a non-health consequence. The following quotes illustrate two other cases where health problems are valued in terms of non-health consequence, in this case social interaction and ability to work:

> *I think because walking is one of my activities, and one of the groups I actually belong to uhm. It would cut off some of my social interaction, so if I had to choose I'd prefer not to have walking problems. (Participant 14)*

> *If you've got moderate problems washing or doing usual activities, and you're below the retirement age, you're going to go on long-term sickness benefit. Finding any work or doing even work from home is going to be a real, a real problem. (Participant 9)*

The distinction between a health domain and the consequences of a health problem can be subtle. For example, some participants mention 'activity limitations' as a consequence, but this cannot be equated with the 'usual activities' domain of the EQ-5D. This is because people used information from all five domains to assess their activity limitations. For example, anxiety or depression was perceived as influencing an individual's ability to do activities:

> *I think uhmm the level of anxiety, and also the level of mobility, walking affect my leisure activity because, uhmmm if I am less anxious I will be more willing to go out. (Participant 1)*

Domains that describe feelings, such as 'pain or discomfort' and 'anxiety or depression', also impact activity-based consequences, while activity-based domains (mobility, usual activities, self-care) were also described as impacting on feelings based consequences. For example, in the quote above anxiety is said to affect activities. In the following quote, walking is thought to affect mood:

> *I actually probably go for walking about, cause then if I can go out, that does elevate, raises me mood. (Participant 3)*

A full list of the consequences mentioned, and the number of times mentioned, are reported in Table 5.5 and Table 5.6. An overview of the number of times these concepts were mentioned by participants may provide an account of the importance of this idea. Of the many consequences mentioned six appeared most frequently. The sixth consequence was mentioned 33 times by 13 participants, while the seventh was mentioned 12 times by seven participants. The six most frequently mentioned consequences were: activities, enjoyment, independence, relationships, dignity, and avoiding being a burden.

Consequence	**Illustrative Quote**
Ability to support others	"Also the choices that you make depend on how it affects those other family members…whether you you still being able to support them"
Achievement	"Your life is more than just existing but achieving something"
Activities	"For me, with, used to being able to do lots of activities and physical stuff etc…It is quite important to me...I've been in situations where I've been ill and it stops you from doing activities you see, and you know it's it's terrible"
Being a burden	"(…) other people would need to sacrifice their time to help me out."
Career and work	"You cannot fulfill or pursue your career or job"
Change in living conditions	"(…) and what changes to housing are gonna be needed, would it mean moving?"
Cognitive functioning	"I just think that, actually can't think properly when you're in a lot of pain...it's very difficult to focus when you're in pain"
Confidence	"(…) these are the things that give me a lot of confidence"
Dignity	"I think washing or dressing is more sort of personal, so for some people it might be harder to have help with that because they feel more embarrassed about that."
Enjoyment	"(…) whether actually your enjoyment of life would be sort of diminished"
Independence and control	"(…) it seems to me that it's very important to be independent"
Income	"I'm keenly aware of the impact that health problems could have on your ability to earn."
Lack of transport	"(…) if you can walk about then…you can still get to things like a hospital and the doctors without having to wait for an ambulance"
Motivation in life	"(…) the less you have the volitional to do the things you were going to do in the first place"
Purpose and contribution	"It helps with uhhm how I feel uhm feeling of a sense of worth and also contributing"
Relationships	"(…) would I be losing friends or losing contact with people?"
Specific time-related goals	"(…) just to have extra years you know seeing my children grow up…compared to life A, whereby you don't have any opportunity to see how they would develop"
Time per day spent on ill health	"Well, I mean that's like an hour of your day, doesn't really matter"

Table 5.5 List of all consequences mentioned by participants, along with example quote

Consequence	Count of participants who mentioned consequence	Count of total mentions
Activities	21	128
Enjoyment	21	100
Independence and control	19	99
Relationships	14	47
Dignity	13	38
Being a burden	12	30
Time per day spent on ill health	6	11
Income	7	11
Ability to support others	4	7
Purpose and contribution	7	6
Lack of transport	5	5
Specific time-related goals	3	5
Career	2	5
Change in living conditions	2	5
Achievement	2	4
Motivation in life	3	3
Confidence	1	2
Cognitive functioning	1	1

Table 5.6 Count of number of participants and total mentions of each consequence, ranked by total number of mentions

Interaction between health domains

When participants estimated the consequences of ill health, they occasionally used multiple EQ-5D-5L domains to determine one consequence. Thus there was an interaction effect between the various EQ-5D-5L domains. For example, the following quote illustrates a participant describing how both 'anxiety or depression' and 'pain or discomforts' interact together:

> *Because this, because severely anxious or depressed can actually make you more, make you experience pain more. (Participant 4)*

In the quote above, problems in one domain were exacerbated by problems in a second domain. Some participants also mentioned the reverse. The following quote illustrates how according to the participant the domains acted as complements and self-care problems were less important given mobility problems:

If I have severe problems in walking about...what am I getting washed or dressed for? So I can, I don't know, sit and watch daytime television? (Participant 9)

These quotes illustrate how multiple domains together can affect the consequences of health states.

Difficulties with estimating consequences

Some participants mentioned difficulties with estimating the consequences of the health problems.

Yeah, it's just it's knowing how that would affect you really, whether actually your enjoyment of life would be sort of diminished. (Participant 5)

Again it's difficult for me to imagine the issues surrounding not being to walk around all the time. (Participant 16)

This meant that participants struggled with how the health state would affect their life, which makes the valuation of the health state more difficult.

Consequences not considered

At the end of the interview, participants were prompted with a list of factors from the literature and were asked whether they had considered them or not. Some participants had not stated some things out loud, although they did think about them:

Work yes, I didn't shout it out loud, this particular one (...) (Participant 15)

Some criteria were not relevant to people. For example, work was not relevant for those that were retired:

I guess I mean I've been retired two and half years now, had you been speaking to me 3 or 4 years ago, I think I would've almost inevitably thought about that much more (Participant 10)

Several people did not consider the consequences because it had not occurred to them. For example, several people forgot the impact of health on their family but did say that they considered it important once mentioned.

Right, I think support, I didn't think of that actually. Support is very very important. Being a main earner at home etcetera, that sort of thing. Doing the DIY. (Participant 12)

Weighing the consequences

Some participants made comments about the importance of the various consequences. Age, personal circumstances, and personal values seemed to be a factor in weighing the consequences:

If I was younger than I am, then I'd always go for the years, but looking at the things that are on that piece of paper, maybe I'd go for less years but better health (Participant 14)

I guess, I don't know if it might be different if I was older, but I'm only 20. So you think living to 30, and putting up with that would be would be worth the trade off (Participant 18)

I think probably for me, because, uhm, I'm married, being, having that independence is is more important (Participant 11)

I mean yes the, I suppose again, personally at all costs I want to maintain a reasonable level of independence for as long as I could. For me personally that would be extremely important. uhm, particularly because I don't want to be a burden on my wife, and because I'm fairly independent person really and I like doing things for me self (Participant 10)

Age and personal circumstances and values seemed to relate to the importance of the consequences for the participants. For example, whether someone is married may relate to what consequences they focused on and how important it was for them.

The theme of consequences displays the extent to which participants focused on the non-health effects of the health state. The non-health effects, which describe what an individual's life would be like given a health state, are an important aspect of valuing health.

5.2.4 Theme 4: task related themes

This last theme describes how the nature of the valuation tools affected the process of valuing health. Some participants made statements about specific aspects of the valuation tasks, such as the health state description system of the EQ-5D-5L or the valuation method used. The sub-themes are described below.

Difficulties in interpretation of attribute levels

Participants mentioned having difficulties with the labels attached to the domains. This was because the labels were seen as being subjective and difficult to interpret:

> *Without any more detail of course you can't really think quite what is meant by some of these statements, because what might be a moderate disability to you might be a slight one or even a non-existent one to me. (Participants 10)*

> *I think we're getting a bit semantic...well I'm thinking about the meaning of, what's the meaning of moderate vs. slight? (Participants 21)*

> *When you say severe difficulties or moderate difficulties, or severe and extreme, what is the difference? (Participants 3)*

Several participants stated that they were unable to differentiate between the attribute levels:

> *So if it was a lot of pain, and it would just be a lot of pain, it wouldn't really be this is severe, oh this is extreme pain. (Participants 17)*

> *Cause to me there isn't any difference. I mean it says slightly anxious or depressed, or moderately anxious or depressed, and to me there isn't any difference. (Participants 11)*

> *You see when yeah when it comes when it comes to physical pain I don't really make distinctions beyond a certain point after moderate, I can't distinguish between severe and extreme. (Participants 13)*

Placement of extreme versus severe

Five participants queried the ranking of extreme compared to severe. Although they did consider extreme as worse than severe, the wording for these two levels of pain or discomfort and anxiety or depression was found to be confusing:

> *Is extreme meant to be worse than severe? (Participants 17)*
>
> *Extreme is more than severe I guess. (Participants 21)*

Coherence and understanding

Some participants discussed their choices in the context of previous choices and commented on how their choice was either consistent or inconsistent with previous choices:

> *So if I were consistent I would be going with life A wouldn't I? (Participant 10)*

There were also statements that indicated that participants had developed or changed their thought process over the course of the task. The following quotes illustrate how participants reconsidered their previous judgements on washing and dressing or mobility:

> *I think, the more you discuss things, and talk about, I mean it's like the problems washing and dressing yourself, which in the first instance seems horrendous, but then when you start talking these kind of things through with somebody else and you think yeah you know probably would adopt that and I could probably adapt to that so it doesn't seem as bad the more you talk about something and the more you thinking about something you think yeah I could probably overcome that uhm, and then you start thinking about what things there is in society, so the more time we stayed thinking about stuff and how you could accommodate it yourself they become less of an issue (Participant 11)*

> *I've become more aware of it in me own mind is slight problems walking about, might not be the end of the world. In that, when I've first came into the room, having problems walking about and I'm sure when you listen back on that you'll see what I'm saying, was, the be all and end all. I need to be mobile, but now I've thought about things, things are changing. (Participant 12)*

Before I couldn't I couldn't separate walking from the usual things, whilst yes I can see that I can do my usual things without necessarily walking yeah. (Participant 21)

There was thus an indication that participants continued to develop their understanding and beliefs throughout the task, while also trying to maintain some consistency with their previous responses.

Difficulties explaining choices

Sometimes participants had problems explaining their choices. This occurred more frequently in the TTO choices:

No I think I might just stick at that point. I don't really know why but yes, I don't know, I don't know why. I can't explain that to be honest. (Participant 10)

Well I'm not quite sure why, but I think I go for life A. (Participant 6)

In the above examples the participant completed the tasks by providing a response but were not able to explain their choice.

This section provided an overview of various factors that were more specific to the valuation task, relating to the various ways in which participants engaged with the task and techniques.

Summary of descriptive account

The first three sections of this chapter covered the three key themes relating to how participants valued health states: the interpretation of a health state, conversion factors, and consequences of health states. The three were described and illustrative quotes were provided.

The next section will attempt to develop an 'explanatory account' of the data, focusing on uncovering links and connections between the three themes. The aim is to explain the process of valuing health from participants reading the health state to expressing their preferences over health states.

5.2.5 An implicit explanatory account

This section will focus on providing an explanatory account of the entire process of health state valuation based on the observations and analysis of the interview data. This section will attempt to identify relationships between the themes that were found in the data and reported in the previous section. The explanatory account developed in this section is implicit. There was little indication of explicit linkages by the participants. For example, a participant may explicitly mention that relationships or support were important to their decision but not explicitly how those factors relate to other themes or how those factors influence their health state preferences:

> *If you've got uhhm, a partner, a wife, husband, partner uhh, who you're very very happy with and who you love, uhhm and you want to be with, I think that makes a great difference. Uhhm, if things were different with me, uhhm, and I was living alone, uhhm, and wasn't in a happy relationship, then perhaps my answers might have been different. (Participant 19)*

However, by reviewing the transcripts patterns emerge about how the themes are linked together. These patterns occur because participants frequently mention certain themes together. From these patterns it is possible to develop an implicit explanatory account.

Three major stages have been identified that can provide an explanation of how participants value health states. The stages are shown in Figure 5.1. It is important to remember that these stages do not imply that participants follow the process in a neat orderly fashion. The process is iterative and the participants move back and forth through the process. Yet the 'stages' of the process can be logically ordered, with each stage logically requiring the previous step. As the stages are displayed in Figure 5.1, they should be seen as an orderly account imposed on a process that in practice does not occur in an orderly fashion. The explanatory account and examples of how the framework was observed in practice are presented in the next three sections.

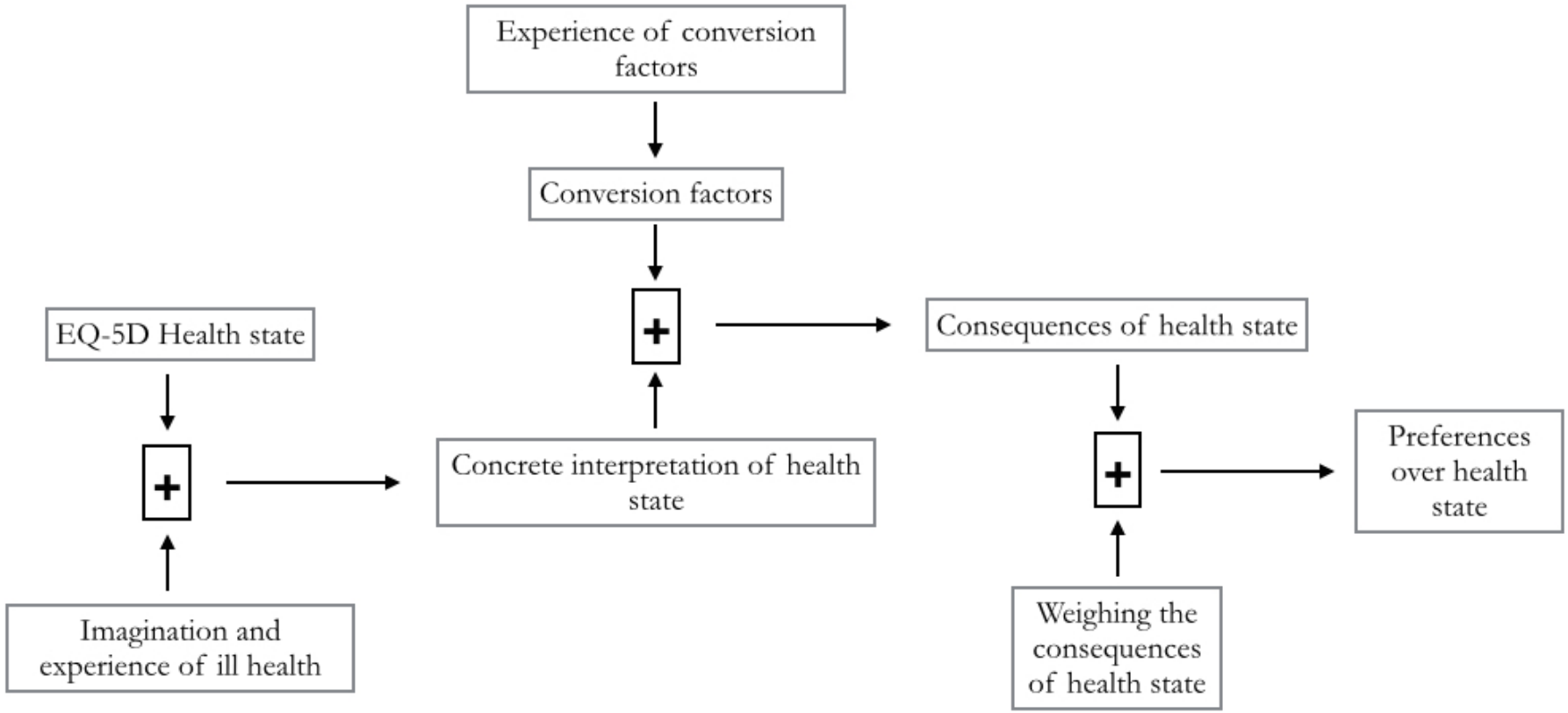


Figure 5.1 Themes and links between themes that explain the process of health state valuation

Link 1: The link between experience with ill health, imagination, and the EQ-5D health state

Upon reading an EQ-5D-5L health state participants often considered what that health state would be like in concrete terms. This was described as concretisation. For example, in the quote below 'moderate' mobility is described more concretely as 'having a walking stick':

> *So you're saying, is an extra 2 years of life worth uhm, moderate problems walking about, uhmmmm, I'm trying to imagine what kind of impact moderate problems walking about would have (...) uhm, moderate problems walking about I'm thinking, if moderate problems does that mean having a walking stick (Participant 11)*

In suggesting a concrete image of the health state, some participants considered previous experience with ill health and others used their imagination:

> *The pain and discomfort in life A is only slight (...) I mean I'm thinking to my self it's a rhetorical questions but (...) what's slight really? (...) I think we all you know, you sort of put your back out, you have slight problems putting your socks on in the morning but (...) it's a nuisance (Participant 10)*

> *So I have moderate pain and discomfort (...) I mean pain, in a sense, I think it's not as consistent it's not as there as anxiety and depression for me (...) but I've never had a condition that has given me chronic pain all the time so I mean, I'm imagining this. (Participant 6)*

In the first above quote specific experiences with of ill health ('put your back out') is used to convert specific health problems ('slight pain or discomfort') to more concrete terms ('problems putting your socks on in the morning'). In the second quote lack of experience meant that the participant used his/her imagination to form a more concrete image of a health problem (i.e. moderate pain or discomfort as a consistent chronic pain).

The comments from the participants suggest that there is a link between the three factors of the EQ-5D health state, imagination, and the experience with ill health. These

three factors together are used to develop a concrete interpretation of the health state. This concrete interpretation is the first stage in valuing health.

Link 2: The link between concrete interpretation of the health state and conversion factors

A link was found between the concrete interpretation of a health state and conversion factors. The data suggest that people consider conversion factors and the concrete interpretation together in order to determine the consequences of health states. In the next quote, not being able to walk about is concretely interpreted as being 'confined to a house'. Those concrete problems combined with the participant's social environment (whether he/she has a partner) leads him/her to become a burden and unhappy:

> *If I can't walk about...then, that will once again confine me to a house (...) and that will make me more depressed (...) it's going to be a personal impact, and it depends on the social scenario a lot. If at this stage I'm partnered with someone, and I'm constantly having problems with walking about, obviously my partner would be very understanding, I would imagine so, but I would feel that I'm imposing a lot on them. So that would make me very upset. (Participant 13)*

In the following quote pain or discomfort and mobility problems are concretely interpreted ('struggle to get sleep' and 'slow, stilted'). They are considered with conversion factors such as medication or aids, and how that determines how the participants will feel:

> *Slight [pain or discomfort] you would be uncomfortable, perhaps struggle to get sleep a little bit, take perhaps some pain killers to make it go away. Whereas severe pain, would be (...) all the time, which I imagine you having it and not necessarily be able to treat it, though you might be able to get some strong drugs, you probably still experience a lot of pain anyway (...) moderate problem walking about (...) would be you [are] perhaps slow, stilted, or you'd need to have an aid (...) I think the constant pain would have more of an effect on you how you felt than not being able to walk about (Participant 20)*

In the next quote, mobility problems are interpreted in the concrete form of 'missing a leg' and being 'frail' and are then linked to the conversion factor of receiving support or using a wheel chair. The participant then considers how that would affect his/her daily ability to complete tasks:

> *I'm trying to think about what's the difference between severe and moderate walking difficulties. I see somebody as severe who is probably missing a leg, or maybe two. That's quite severe in terms of walking about. But moderate could be, you know, they're very frail or can't walk particularly well. You know both of them, pretty much end up with the same situation of needing a wheel chair or some form of assistance. Uhm, so that's not much of a consideration as actually being able to do your day to day tasks. (Participant 15)*

In the second stage of the explanatory account the concretised health problems and the conversion factors helped the participant decide on an effect of the health state on them personally. The example quotes show how when someone considered the health state and has added conversion factors, they determine the impact that health state would have, for example whether it would result in them being a burden, or feeling worse. Combining the concrete interpretation and the conversion factors thus results in a list of consequences of that health state.

Link 3: Link between the consequences, weighing the consequences, and preferences

A link was found between the consequences and preferences. This link was mediated by consideration of the importance of various consequences. The quote below illustrates a participant weighing different consequences:

> *Uhm I, I, definitely like to travel a lot but if it...if it came down to limited life I'd I'd find the people in my life more important, so that would definitely be a bigger issue. (Participant 18)*

It is when participants have weighed these consequences that they ultimately come to make choices between health states. For example in the quote below the participant

explains the importance of ability to support others and career prospects and how it influenced her health state valuations:

> *I think work isn't the most important thing in my life it (...) I think family are. You know work is something that you definitely have to think of, but I don't have a career and I'm not about career progression (...) So it wouldn't be the first thing that would come to my priority, it would be more about family (...) the choices that you make depend on how it affects those other family members, let's say for example I, if you've got a family member that relies on you quite a lot, then that's something that's gonna be on the forefront of your mind, of like whether you you still being able to support them I think. (Participant 4)*

Full Examples of the explanatory account

Figure 5.1 shows the entire process of health state valuation incorporating the three stages outlined in the explanatory account. Keeping in mind that this was an implicit explanatory account, the following quote illustrates a more explicit case of this framework in a participant's statement during health state valuation. The participant starts with a description of the health state. S/he considers what it would mean concretely and how the relevant conversion factors would determine the consequences of the health state. S/he then weighs those consequences. Finally, s/he expresses his/her preference (each stage is noted in brackets):

I'm thinking severe problems doing your usual activities [health state description], I'm trying to think of, so I couldn't do anything, I couldn't come to work, I couldn't do any hobbies, I couldn't do anything with the children [concretisation], and then my immediate thoughts are trying to think of what support networks have I got in place [conversion factor]. Could I still have a, a fairly good quality of life [consequences] and what is the support networks [conversion factor] and is that worth two years [weighing consequences]. And I'm thinking immediately of, drawing on family, and friends, and then possibly into a wider circle, in terms of, what the uhm, NHS provide in terms of support and what probably charities linked to whatever kind of medical condition I had could provide [conversion factor] uhm. Then the severe pain or discomfort, I'm thinking is severe pain or discomfort, all the time [concretisation]? Or can that be stopped with uhm pain killers and pain relief [conversion factors], uhm, yeah and then I'm thinking of, our friend's mum who I think has [disease name removed] which (...) is supposed to be quite painful and she's basically housebound [experience ill health] but I've never heard of them saying that, you know, she's had enough or that she's suicidal [concretisation] so based on that experience I would hmm.....hmm I don't know. I'm just thinking, two years, severe problems, you got 8 years of full quality of life compared to 10 years hmm...the 8 years you see sounds a long time of being able to do everything, and then the two years of not being able to do. I would probably go for life A on that one [preference expressed] (Participant 11)

In the above quote the participant starts with a health state and considers what severe problems in usual activities would mean to her life. She then considers her support network and whether that allows her to maintain a good 'quality of life'. The same occurs with severe 'pain or discomfort', where s/he mentions the experience of her friends, who is housebound and in pain. Finally, having considered all these factors s/he expresses her preferences.

The stages of the process are not completed in an ordered linear fashion. In fact, participants jump backwards and forwards in considering various factors. This jumping back and forth is to be expected in think aloud exercises, which are concurrent verbalisation of thoughts and not ordered retrospective justifications. It is not suggested that the participants complete the framework in a mechanical straightforward fashion, but the stages are distinct and pre-suppose each other. For example, to assess how the individual's environment will interact with 'moderate mobility' problems, the individual has to decide what the concrete meaning of ‘moderate mobility’ problems is.

In the next and final section of the result section, the explanatory account is critically reviewed.

5.2.6 Critical review of the explanatory account

In this section the three techniques by Miles and Huberman (1994) are applied to the explanatory account. The techniques are to look for outliers, to look for negative evidence, and to find and test rival explanations.

Looking for outliers

In assessing the explanatory account of Figure 5.1, one technique is to look for outliers. An outlier would be a participant that did not discuss consequences, conversion factors or did not concretise the health state. In the data no outliers were found. All participants used a similar process of valuing health. The framework charts were reviewed (charts not shown here) and there were no cases where a participant did not consider either the consequences of a health state, conversion factors, or concretisation of a health state. One participant had not mentioned any experience or use of their imagination when concretising the health state but the other stages were present.

Look for negative evidence

The transcripts and framework charts were reviewed to find instances of direct negative evidence. One finding that could provide negative evidence of the explanatory account is the theme of difficulty in explaining choices (see page 100). It was discussed how several participants noted that they had difficulty explaining their choices. This can be considered a contradiction to the explanatory account because the explanatory accounts attempts to explain choices that participants considered unexplainable. Overall, this theme was not frequently encountered. It was mentioned only three times, two times during the TTO and one times during the DCE. In the TTO the unexplained choices referred to explaining the precise indifference point of the TTO, rather than the entire

TTO process. For example, it may have been difficult to explain why the precise indifference point was at eight years but not for why it was approximately eight years. Overall there was no overwhelming negative evidence found.

Find and test rival explanations

Two rival explanations are available in the health economics literature to explain the health state preferences. First, Murray et al. (2002, p.736) argue that health can be measured qua health and that health state valuations are "the reductions in health associated with particular health states", and they "define health state valuations to be simply indices of overall levels of health." This theory argues that health state values are a uni-dimensional measure of health, rather than the "desirability of a life given the health state" or "goodness" of health (Murray et al., 2002, p.737). It would imply that no non-health considerations would be important in the process of valuing health. The theory contradicts the explanatory account presented in this chapter because the account it includes non-health considerations such as conversion factors and consequences. The data in this study do not conform to the theory that health state values are health indices. Participants frequently mention non-health factors and that suggest that they are necessary to value health. No participant valued the states in a way that is compatible with valuation as an index of health. For example, no one claimed that in the DCE, one choice represents simply 'more health' or was 'healthier' than another choice, without any need to consider non-health factors.

A second theory is that a part of people's preferences are affective reactions to the health state. For example, preferences may be formed by the "initial shock reaction to, or fear associated with, that state" (Dolan and Kahneman, 2007). This theory is not inconsistent with the findings in this chapter because the affective reactions would likely be based on individuals' interpretations and perceived consequences of the health state. Therefore, this theory would not necessarily invalidate the framework provided in this

chapter, but may suggest that it is incomplete. It would for example suggest that the participants' assessments of the consequences may be overly pessimistic because the assessment is based on shock reactions to the health state. The findings of this study would suggest that individual's health state valuations are not entirely based on their affective reaction because individuals do provide reasons for their valuations.

Overall, the first rival theory is not well supported by the evidence in this chapter and the second would not invalidate the explanation provided.

5.3 Discussion

5.3.1 Summary

The think-aloud evidence presented in this chapter suggests that preferences over health states are non-basic because they depend on beliefs. Preferences over health states are not just matters of taste. Participants go through several stages while valuing health states. The EQ-5D state description is combined with participants' experience and imagination to form a concretised health state. The concretised state is combined with conversion factors to determine the consequences of the health state. Participants in this study most frequently mentioned six consequences: activities, enjoyment, independence, relationships, dignity, and avoiding being a burden. The consequences were then weighed to come to an overall evaluation of health states. The role of the non-health consequences of the health state means that the utility value of a health state is a measure of "the desirability of life given a particular state of health" (Tsevat, 2000) and not a measure of health. Preferences over health state depend on complex cognitive evaluations which depend on people's beliefs about the health state and the interaction of the health state and their personal and social situation.

5.3.2 Comparison to literature

Other qualitative studies on the process of health state valuation have found that participants consider many non-health factors, such as religious beliefs, available support, age, and the achievement of specific goals (Robinson et al., 1997; Baker and Robinson, 2004; Osch and Stiggelbout, 2007). Health and non-health cannot be seen as separate factors when valuing health states. To be separable would mean that health could be evaluated without considering any non-health aspects, but as Broome (2002) argues, "asthma is less bad if you are well housed, mental handicap less bad in supportive communities, blindness less bad if you have access to the internet...The interaction between health and other features of a person's life is so intimate that health cannot be treated as separable." This study expands on the notion that non-health factors are important by describing the relationship between the health and non-health factors, and emphasising how non-heath factors influence health state values. A distinction is made between two types of non-health factor. One set, which were named 'conversion factors', are aspects of an individual's personal and social situation that interact with the health state. The second set, the consequences, are the final effect of the health state on an individual's life.

The six most frequently mentioned consequences (activities, enjoyment, independence, relationships, dignity, and avoiding being a burden) are supported by previous research. Although the consequences may appear foreign to health economists focused on income and leisure (Sendi and Brouwer, 2005; Krol et al., 2009; Tilling et al., 2010b), income was mentioned but seen as a means to an end and leisure is subsumed in other consequences such as personal relationships and activities. Table 5.7 displays the domains of the ICECAP-A questionnaires, which measure quality of life in the general public (Coast et al., 2008; Al-Janabi et al., 2012), a review of quality of life domains relevant to mental health (Connell et al., 2014), philosophical account of five

important domains for humans provide by Griffin (1986, p.96), and suggestion on how health states can be compared from Hausman (2012a, p.122). There is a clear overlap between the findings in this study and these four lists. This would indicate that the consequences are generalizable to a larger sample than in this study.

Top six consequences of this study	**ICECAP-A**[1]	**Mental health QoL**[2]	**Griffin**[3]	**Hausman**[4]
Activities	Achievement	Activity	Achievement	Occupations and personal projects
Enjoyment	Enjoyment	Well-being	Enjoyment	Feelings
Independence	Autonomy	Autonomy, control and choice	The components of human existence: agency, autonomy, basic capacities, and liberty	The ability to live independently and make one's own choices
Relationships	Attachment	Relationships and a sense of belonging	Deep personal relations	Recreations, social interactions, or family life
Dignity	-	Self-perception	-	-
Being a burden	-	-	-	Burden on others
-	Stability	-	-	-
-	-	-	Understanding	-
-	-	Hope and hopelessness	-	-

1: (Al-Janabi et al., 2012), 2: (Connell et al., 2014), 3: (Griffin, 1986), 4: (Hausman, 2012b, p.123)

Table 5.7 Similarity between empirical and philosophical perspectives on quality of life and top six frequently mentioned consequences by participants in this study

5.3.3 Implications for practice

The results of this study have implications for the type of descriptive systems used in valuing health, the design of the valuation tasks, and the use of sub-group preference values used in cost-effectiveness studies. The fact that the six consequences appear be

important in valuing health may have implications for how the health states should be described. The final six consequences appear closer to the ICECAP-A domains than the EQ-5D domains. This may be an argument for directly using the ICECAP-A, a broad 'well-being' measure, in economic evaluation. A disadvantage of using a broad measure would be the possible lack of responsiveness or sensitivity in such a measure when dealing with health interventions (Mukuria et al., 2015). It may be that the ability of members of the general public to adequately assess the non-health consequences of health has to be traded off with the responsiveness of a health-related questionnaire to determine whether a well-being or health measure is preferred.

The results have implications for the design of health state valuation tasks. First some minor suggestions can be made. Participants struggled with the placement of the extreme and severe labels used in the EQ-5D-5L. Participants could perhaps be guided to more clarity here by making the distinction between the levels clearer. Some participants attempted to re-write the health state, in particular when a state was deemed unrealistic. Indeed the psychology literature has previously identified a tendency amongst participants for "restructuring the decision problem to create dominance and thus reduce conflict and indecision" (Slovic, 1995, p.369). Care should be taken to check the face validity of health state descriptions in valuation exercises and perhaps unrealistic states should not be included at all. A possible bias observed was that participants considered treatments while valuing the health state. Even though participants were instructed to "imagine that you would be in this health state even if you were receiving treatment", they often continued to consider treatment possibilities. Some instructions when valuing health may have to be re-considered. For example, Feeny et al. (2002) instruct respondents that "when imaging yourself in these health states please remember that where you live, your income, your friends, and family would be the same as now" in order to "focus only on differences in health state

descriptions." The findings of this study suggest that this would be impossible for participants, and that it is precisely changes in the non-health factors that participants are valuing.

Health state valuation methods could perhaps be improved by allowing participants to reflect and deliberate on the consequences of health sates. This study has shown that many beliefs are involved in constructing preferences over health states. For example, if two people's experiences with mobility problems are different, their concrete interpretation will also be different. This may cause them to make different judgements about the consequences of that health states, and thus their final preferences will likely be different, despite not having inherently different preferences. Participants have to make judgements about the consequences of a health state and the TTO and DCE tasks require participants to make these judgements during the tasks. Some participants expressed doubt and reservations about estimating these consequences. In addition, participants expressed a dynamic understanding of the task and their ideas were changing as they valued more states. All of this suggests that participants may benefit from a more structured approach that allows them to consider the consequences of health states directly and allows participants time to reflect and deliberate on the consequences before they make judgements on the (un)desirability of different health states.

The implications of the non-separability of health and non-health factors may mean subgroup values should be used in cost-effectiveness studies (Gold et al., 1996, p.103). For example, Flynn et al. (2010) finds that those living alone and those with a partner have different health state preferences. The differences can be explained either by differences in conversion factors (for example, more available support) or by differences in weighing of consequences (for example, caring more about independence). There are counter arguments for using sub-group preferences

(Robinson and Parkin, 2002). For example, the practicality of using sub-groups could be difficult because of issues such as identifying preference sub-groups, determining when preferences are sufficiently similar and dissimilar to declare a group a sub-group, and determining which characteristics could be (morally) considered to distinguish sub-groups (Sculpher and Gafni, 2001). There is also an argument that sub-group preferences have not been proven to lead to higher allocative efficiency and may in fact reduce efficiency (Robinson and Parkin, 2002). Robinson and Parkin (2002) argue that if an intervention is provided to a sub-group "there is an opportunity cost in terms of other health care forgone. From the point of view of the community as a whole, the other health care forgone will be valued more highly than the treatment provided to the sub-group. Efficiency will therefore be reduced." The reason for their comparison is not clear because the obtained or forgone health care is not necessarily for the entire community. Indeed, later Robinson and Parkin (2002) suggest that "efficiency will be achieved using the preferences of sub-groups…where there are separate allocative decisions to be made for each group". Robinson and Parkin (2002) then suggest that this requires either different health budgets or different ICER thresholds from one common budget and that the "justification of the different budgets or ICER thresholds would be – in terms of efficiency, equity, need, social justice or whatever – is unclear." Indeed the assumption may have to be made that the difference between dead and full health is the same for each sub-group to maintain the same ICER threshold. There are thus practical and equity issues surrounding the use of sub-group preferences.

5.3.4 Limitations of study

One limitation of this study is the use of the think-aloud protocol as the method for investigating participants' thought processes. Participants may have not fully verbalised their thoughts. Wildman (1995) argues that what participants say may be

unrepresentative of their actual thoughts, for example they may censor negative opinions. In addition, there may be subconscious processes that affect participant's responses to health state valuation tasks but that are not verbalised. For example, it is known that the iterative procedure can influence TTO valuations (Lenert et al., 1998) but the framework reported in this chapter does not explain this phenomenon. There may also be a concern that thinking aloud will influence or even change a participant's thought process. However it has been found that when comparing the performance of a control group and a think-aloud group the think-aloud group was slower but their performance was not changed (Ericsson and Simon, 1993, pp.84-89; Hertzum et al., 2009). These limitations cannot be avoided in the think-aloud protocol. They would mean that the explanatory account is incomplete, but not necessarily wrong.

The sample in this study was non-representative and homogenous. The majority of interviewees were identified via the university mailing lists and only those with access to email would be able to receive the invitation email. In particular, the sample was more educated than the UK population, contained more students, less retirees, and is younger (Office for National Statistics, 2011). Because of these limitations and the small sample size, it is difficult to make claims on the statistical representation of the themes that emerged. However, increasing the sample size would likely not have led to an increase in the number of themes found. No new themes were identified in the last three interviews indicating that data saturation was reached and within the recruitment sources used the sample size may have been adequate (Ritchie et al., 2003a, p.80).

Six consequences were identified as core aspect of valuing health in this study by counting the number of mentions for each consequence. Counting qualitative data assumes that the frequency of mentions reflects their importance but this does not have to be the case. The top six were taken as most important because there is a fairly large gap between number six ('being a burden') and seven ('time per day spent on ill health').

The sixth consequence was mentioned 33 times by 13 participants, while the seventh was mentioned 12 times by seven participants. Many of the less frequently mentioned consequences could be defined as aspects of the six most frequent consequences. For example, 'lack of transport' may ultimately mean lack of activities or enjoyment. The six consequences may not be reflective of the larger UK population because the sample was non-representative. The similarity of the six consequences to domains of the ICECAP-A and quality of life domains for mental health provides some indication of the validity of the findings, but care should be taken when interpreting the results.

The findings of this study may not be generalizable to other settings, such as for example other instruments. It is possible that a different instrument, such as the SF-6D, or a different method, such as the SG and VAS, would cause participants to consider different things. There is both qualitative and quantitative evidence indeed that VAS is different from the TTO (Robinson et al., 1997). In this study the EQ-5D was chosen because it is the recommended tool in the NICE reference. The DCE and TTO were chosen because they are used to value the EQ-5D-5L (Oppe et al., 2014).

5.3.5 Future research

Three suggestions for future research can be made. First, the use of reflection and deliberation to help individuals value health states can be explored. The importance of consequences of health states in valuing health suggests that reflection and deliberation could focus on the consequences of ill health. This direction of research is explored in Chapter 6. Second, the ability of members of the public to judge the consequences of health states correctly can be taken as an indication of how informed they are about health states. Some participants expressed doubt about whether they were able to do so and some participants expressed doubt about other people's ability to do so. This research could help establish whether participants are informed about what they are

valuing. This direction of research is explored in Chapter 8. Third, the explanatory account reported in this chapter could be tested quantitatively. For example, some of the non-health consequences or conversion factors could be specified along with the health state to see whether the health state values will change. This could be similar to recent work by Dolan et al. (2013), who added life satisfaction to health state descriptions to see if valuations changed.

5.4 Conclusion

One of the assumptions considered in this thesis was whether preferences over health states are purely matters of tastes. The findings of this study would indicate that preferences over health states are not matters of pure tastes or gut instincts and therefore should not be treated as such. The first objective of this thesis was to describe the process of how individuals value health states. In this study the participants considered a variety of beliefs when valuing health states. Ultimately, participants evaluated health states by judging severity and importance of the non-health consequences of ill health for their life.

The findings suggest that valuing health is a cognitively complex and subjective task. Participants must consider the practical implications of an abstract state, how the health state relates to their environment and personal circumstances, make accurate estimates of the consequences of ill health on their life, and weigh up those consequences. Participants do not appear to come to the task with a set of pre-determined preferences but have to reflect on their environment, personal values, personal circumstances and experiences. Indeed, participants sometimes express that their thoughts and beliefs changed over the course of the valuation task. In order to simplify the task for participant, providing them with a structured task around the consequences may be beneficial because it can help people develop their preferences. In addition, establishing

whether members of the general public are informed about the consequences of health states may be useful to establish the validity of the method used to value health.

The next chapter describes a reflective and deliberative exercise that explicitly considers the consequences of a health state and tests whether reflection and deliberation on these consequences change individuals' health state preferences.

Chapter 6

The effect of reflection and deliberation on health state values

Health states are conventionally valued using one-off interviews that may not give participants enough time to reflect on the complex process of valuing health revealed in the interviews in the last chapter, nor any opportunity to discuss their views with others. That may not be a problem if participants arrive at interviews with a set of complete preferences but the assumption of completeness may not be appropriate. If participants construct their preferences during the task, then preferences elicited from conventional one-off interview methods may be not well-constructed and not adequate to be used for public policy.

One method advocated to help participants construct their preferences is the use of reflection and deliberation. As noted in Chapter 2, reflection is defined as "the action or process of thinking carefully or deeply about a particular subject, typically involving influence from one's past life and experiences" (Oxford University Press, 2015) and deliberation is defined as "a discussion and consideration by a group of persons" (Merriam-Webster, 2015). Giving members of the general public more time to reflect allows them to think about the various conversion factors and consequences of the health state, which were considered by participants in study one (reported in Chapter 5) when valuing health. Giving members of the general public an opportunity to deliberate allows them to make use of experience and knowledge of other people, which is important because in the study one previous experience of ill health and knowledge of conversion factors was seen to be relevant in valuing health.

In this chapter the impact of deliberation and reflection on preferences of members of the general public over health states is studied. The aim is to understand if providing members of the public with an opportunity to engage in reflection and deliberation will change their health state preferences and whether these changes are practically important enough for policy makers to consider because they may have a significant effect on cost-effectiveness rankings of interventions. If so, conventional methods may have to be adjusted so that participants can be given more guidance to develop their preferences during the valuation task. In this chapter only the quantitative effect of reflection and deliberation is discussed. In the next chapter (Chapter 7), a mixed methods approach is used to attempt to explain the quantitative effect of reflection and deliberation.

This study differs from existing deliberative or reflective exercises in two key ways. First, an attempt is made to structure the exercise using the findings from study one (reported in Chapter 5), where it was argued that health states are valued according to their consequences. To do so it uses a MCDA exercise that has not been previously used. The use of MCDA allows the reflection and deliberation exercise to be formally structured. This study also uses the TTO and EQ-5D. The effect of reflection and deliberation has not been previously investigated using the TTO and EQ-5D. This study should provide evidence about whether the findings in the literature on reflection and deliberation can be generalised to the TTO and EQ-5D.

The methods are explained below, including a description of the overall design of the study, the design of the group meetings, the MCDA method, the TTO method, and the recruitment method. Then the methods of analysis are explained. After the methods section, the results are presented. Finally, the findings are discussed, with a focus on the potential role of deliberation and reflection exercises in developing complete preferences over health states.

6.1 Methods

The design of this study is a pretest-posttest design with no control group (Cohen et al., 2007, p.282). Health states were valued using the TTO before and after a group-based reflection and deliberation exercise. Several group meetings were held but each participant was allocated to only one group meeting. In the group meeting each participant completed TTO tasks, a MCDA task, a deliberative task, and then the TTO tasks again. The effect of the reflection and deliberation tasks is measured by comparing the TTO valuations of the participants before and after the group meetings.

The primary justification for the use of group meetings is to implement deliberation. Although reflection can be conducted individually, deliberation requires the exchange of ideas with others. Deliberation could be implemented by various methods, for example by requesting participants to discuss their answers at home with relatives or friends, but there is no guarantee that participants would engage in deliberation and the researcher would not be aware of the content of the deliberation. Qualitative analysis is important to understand the reasons for any changes to health state preferences and therefore it is necessary to be aware of the content of the deliberation. The easiest way to ensure that deliberation takes place and to be aware of its content is to include a group meeting as part of the study design.

In the next section the design of each group meeting is reported. This is followed by a description of how the health states were selected, the design of the TTO and MCDA exercises, recruitment methods and finally, the analysis method.

The ScHARR Ethics Committee approved the study (see Appendix 9).

6.1.1 The design of the group meeting

The design of the group meeting is based on the findings of study one in Chapter 5, where it was found that people's health state values were based on a variety of beliefs

but ultimately on their beliefs about the way a health state would impact on their life. A pilot was conducted to test the protocol. The next few subsections describe the protocol as it was actually implemented, but this section also describes the original pilot and how the pilot informed the final design.

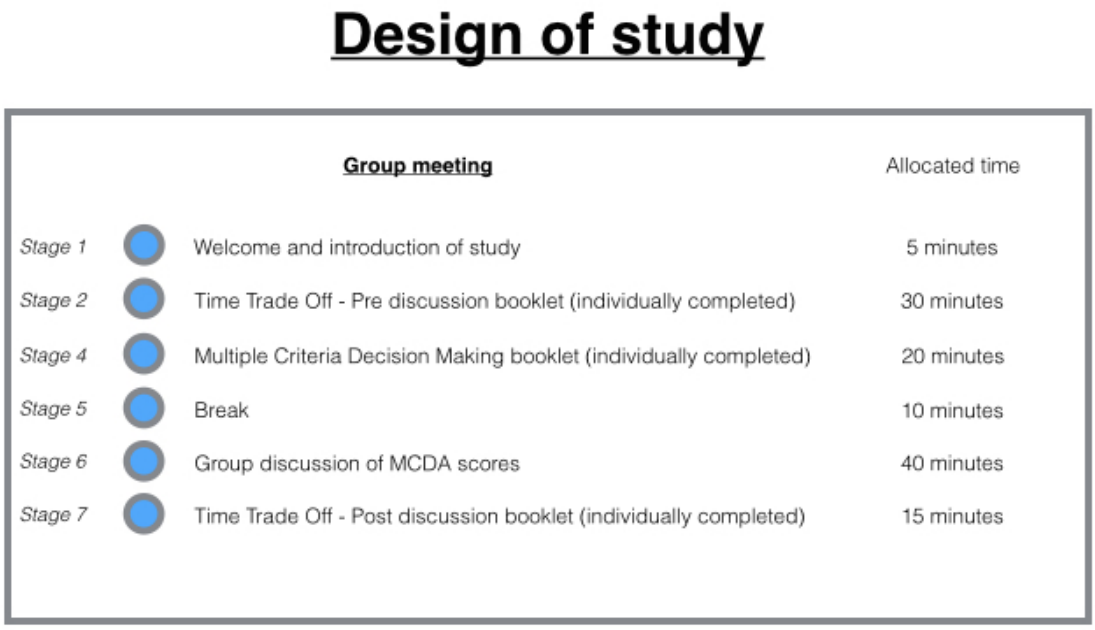


Figure 6.1 Stages of group meeting

Stages of a group meeting

Each group meeting contained several stages: the introduction, the first TTO booklet, the MCDA booklet, a break, the discussion of the MCDA booklet, and the second TTO. The design is reported in Figure 6.1. All participants individually completed the TTO and MCDA booklets but the discussion of the MCDA booklet was done as a group. The introduction and two booklets took about one hour to be completed. After the two booklets were completed, participants had a 10 minute break, followed by a group discussion of 40 minutes. Finally, about 15 minutes was left for individual completion of the second TTO booklet. The entire group process lasted about two hours.

All group meetings were held on the campus of the University of Sheffield. Nametags were given to all participants because it gave participants the ability to

respond to each other's comments more easily. Participants were given information sheets and were invited to sign the consent forms. Besides the participants, only the facilitator (the PhD student) was present during the group discussions. After all the consent forms were signed, the study background and design was explained to the participants.

The TTO was used to value the health states. Although DCE was used in the last chapter, it was not used here. DCE requires estimating models to establish the value of a health state (Brazier et al., 2007, p.164). Participants would have to value many more health states in the DCE than in the TTO to estimate values for each health state. Because of the lower burden of obtaining health state values using the TTO compared to the DCE, the TTO was a more practical choice for this study. The TTO booklet is described in section 6.1.3.

MCDA was used to structure the reflection and deliberation task. MCDA is an "umbrella term to describe a collection of formal approaches which seek to take explicit account of multiple criteria in helping individuals or groups explore decisions that matter" (Belton and Stewart, 2001, p.2). MCDA allows the group meeting to be structured around formal pre-determined steps and may make the group meeting more productive than an unstructured group meeting. MCDA also attempts to guide participants to develop their preferences over health states from more general values, which is more in line with the preference construction literature (Fischhoff, 1991). The MCDA booklet is described in section 6.1.4.

After the participants completed the MCDA booklet, the participants had a break. A break was necessary as both the TTO and MCDA booklet were cognitively burdensome. Cold drinks and refreshments were offered during the break.

After the break, participants were given instructions for the deliberation section of the group meeting. The facilitator emphasised that the aim was not for participants to

agree with each other or to reach consensus. This was emphasised because the objective of this study was to look at whether reflection and deliberation changed individuals' preferences rather than whether consensus values are different from mean values. During the deliberation participants discussed their responses to the MCDA booklet. The deliberation focused on the participants' beliefs on the consequences of the different health states. The facilitator attempted to allow a free discussion while maintaining a focus on the MCDA booklet in order to limit the amount of off-topic discussion.

After the deliberation participants were given the final TTO booklet and a reminder on how to complete the TTO. Depending on the time left, a small debrief was conducted with the group. During the debrief period participants asked questions on the aim of the study and discussed their feelings about the study. The topic guide used by the facilitator is reproduced in Appendix 10.

Facilitator role

The role of the facilitator during group meetings is important (Morgan, 1996; Morrison-Beedy et al., 2001). The facilitator must encourage the group to complete the stages of the meeting while minimising bias due to his/her actions. Bias can occur if, for example, the facilitator emphasises or expresses agreement with a comment of a participant. In addition, the role of the facilitator was to explain the TTO and the MCDA methods. The same facilitator (the PhD student) was present for each group meeting and thus there should be little inter-group variation due to the moderator.

Piloting

Pilot testing was conducted before the above design of the group meetings were finalised. The TTO and MCDA booklets were first piloted with two individuals separately (not in a group setting). This avoided the need to recruit an entire group. Small modifications were made to both booklets based on feedback by the participants.

Some health states were changed because they were found to be unfeasible in both pilots and minor formatting changes were made.

Next a group pilot was conducted. A convenience sample of six students was recruited via email. ScHARR students familiar with the TTO and health state valuation were not recruited. The participants first completed the TTO booklet. After this two participants completed two tasks. The first was a poster task and the second was the MCDA task.

First, a poster was put on the table. The idea of the poster was based on the group method of 'Deliberative mapping', where posters are used to enable participants to express and share their views with in the group meeting (Burgess et al., 2007). The poster listed the health states on the left hand side and had a scale going from 'Not at all' to 'Completely' on the top. Above the scale, one of the six criteria was listed.[14] Each participant was given a sticker and asked to put their sticker on the scale, thus expressing their view on how bad each health state was. After this, the group discussed why they had put the stickers in that way and participants justified their position. This was repeated with the remaining five criteria. For the second task, the participants were then given a MCDA booklet to complete (this MCDA task is described in detail in Section 6.1.4).

After these tasks, a break was scheduled in order to not overwhelm the participants. Following the break there was a discussion around differences between MCDA and TTO tasks and participant's general perception of the impact of health states on the criteria. This was followed by a second TTO.

After this group pilot several changes were made. Based on participants' comments, slight modifications were made to the booklets and more instructions were added. More

14. For a discussion on the criteria see Section 6.1.4

significantly, it was felt that there were too many tasks to complete, which meant little time to concentrate on each task or to have an in-depth discussion. It was also felt that if the majority of the two hours were filled with individual tasks that participants would get tired and overloaded. It was decided that more time should be spent on deliberation rather than the individual tasks. It was felt that the MCDA exercise was critical to get participants to reflect and deliberate on the consequences of the health state. Therefore, the MCDA exercise was chosen as the main tool to keep during the group meeting and the poster task was dropped. This would save significant time and would allow more time for deliberation. This new format was easier to manage and allowed more time for in depth discussion. Furthermore there was less chance of fatigue with fewer stages in the group. This design was chosen for the main study.

6.1.2 Selection of heath states to be valued

Participants valued six health states. The number of states was chosen such that the entire group meeting could be completed within two hours. Based on previous studies (Gudex, 1994, p.29) and the pilot of this study, it was decided that six health states would be the highest number of states that could be feasibly valued and discussed in a group meeting lasting about two hours.

Selecting health states is to a degree arbitrary and no clear guidelines are available in the literature on how to select states. The six health states were chosen to cover a range of severity and be realistic. States with different severities were chosen because the change of health state values after reflection and deliberation can be associated with the severity of health states. It is possible, for example, that valuing severe states is cognitively complex because it may require considering problems in several domains simultaneously or that individuals are not likely to have previously experienced severe health states. In that case more severe states may involve more preference construction

and may benefit more from reflection and deliberation. Indeed some studies found that a group exercise resulted in more changes to the values of states with greater severity than the values of those health states with lesser severity (Krabbe et al., 1996; McIntosh et al., 2007). Within the range of severity the health states were chosen so that participant valued deficits in all domains of the EQ-5D

Health states were also chosen to avoid unrealistic health states. Unrealistic states are states that lack face validity for the participants because the participants do not consider them to be possible in real life. For example, a health state where one is unable to walk about but has no problems in usual activities may not be realistic. Participants in the first study found it difficult to value unrealistic states and frequently re-wrote those states. The participants in the pilot of this study had similar difficulties. Avoiding unrealistic states also meant that the deliberation period would avoid discussion on the realism of the states, which was not the aim of this study. One way to determine which states may be realistic is to determine whether patients report being in those states. If patients frequently report a state it can be assumed to be realistic but if patients do not frequently report a state the health state may be unrealistic. Janssen et al. (2013) report on a study where over 3900 participants, including eight patient groups, in six countries completed the EQ-5D-5L (the full dataset is not published online but was obtained by private communication with lead author M.F. Janssen, 2013). When selecting states for this study, preference was given to states that were reported relatively frequently in that dataset. From the selected states, five of the six states were in the 70^{th} percentile of most reported health states. The six health states are listed in Table 6.1.

State	Description	Arbitrary chosen label[1]
44535	I have severe problems in walking about I have severe problems washing or dressing myself I am unable to do my usual activities I have moderate pain or discomfort I am extremely anxious or depressed	GY
11331	I have no problems in walking about I have no problems washing or dressing myself I have moderate problems doing my usual activities I have moderate pain or discomfort I am not anxious or depressed	YM
32322	I have moderate problems in walking about I have slight problems washing or dressing myself I have moderate problems doing my usual activities I have slight pain or discomfort I am slightly anxious or depressed	IR
31131	I have moderate problems in walking about I have no problems washing or dressing myself I have no problems doing my usual activities I have moderate pain or discomfort I am not anxious or depressed	YC
11334	I have no problems in walking about I have no problems washing or dressing myself I have moderate problems doing my usual activities I have moderate pain or discomfort I am severely anxious or depressed	UI
44553	I have severe problems in walking about I have severe problems washing or dressing myself I am unable to do my usual activities I have extreme pain or discomfort I am moderately anxious or depressed	NA

1: An arbitrary label was provided for each health state so that participants could refer to a health state during the deliberation period without reading out the whole health state

Table 6.1 Health states used in study two

6.1.3 The time trade off task

The self-completion method based on the version developed by the Measurement and Valuation of Health (MVH) project was selected because this had been thoroughly tested in previous research and found to be valid and reliable (Gudex, 1994; Dolan et al., 1996b). It had to be feasible for the before and after TTO tasks to be administered

by one moderator to multiple participants in the group. This requirement meant that a method based on 'props' was not feasible because props require one interviewer per participant (Gudex, 1994). Computer based methods were also not feasible because providing a computer for each participant was not practical and some participants may not be comfortable with technology. Although recent studies valuing the EQ-5D-5L use the lead-time method, it has not been implemented using a self-complete format using pen and paper and preference was given to using a method that was more established using the self-complete pen and paper format.

The design of the TTO was top down titration with a 20 year time span. 20 years was chosen as it allows more precise valuation than the conventional 10 years because smaller units of time can be traded. In the top row of each page, the participant indicates whether they prefer 'Choice A' (full health), 'Choice B', or were indifferent between the two. In the rows below the number of years in full health is reduced by 6 months increments. If the participant prefers no years in Choice A (i.e. being dead) over twenty years in Choice B, the state is considered worse than dead. For states worse than dead, participants compared a combination of ill health followed by full health to being dead. The point of indifference is taken as the average number of years if multiple indifference points are indicated (Gudex, 1994). If no point of indifference is indicated the midpoint between when the participant switches between Choice A and B is taken as the indifference point (Gudex, 1994).

Two self-completion TTO booklets were prepared, one for before and one for after the group meeting. The two booklets were identical except the first contained a form for participants' own EQ-5D-5L, a ranking exercise, and a practice TTO and the second TTO included a one page background form at the end of the booklet. The ranking exercise of the first TTO booklet asked participants to rank the six health states along with the states 'dead' and 11111. The practice TTO in the first TTO booklet used health

state 22222, which was not one of the six in the main valuation. The order of the six states was randomised before designing the booklet. Once randomised, the order was initially designed to be the same for all participants and both booklets. The order of two states was reversed after six group meetings because a high level of inconsistency was noted after the first six group meetings (this is discussed in section 6.3.5).

The booklets are shown in Appendix 11 and Appendix 12.

6.1.4 The MCDA exercise

The MCDA booklet is the second booklet the participants completed individually. An MCDA exercise generally has four steps (Belton and Stewart, 2001):

1. Selecting criteria
2. Scoring options on each criterion
3. Weighing criteria
4. Aggregate scores and weights to calculate overall score

Only the first three steps were conducted in each group meeting. In this study, MCDA is used to structure the reflective and deliberative exercise. There was little or no time in the group meeting for calculating an overall score and provide the overall scores back to participants.

The MCDA was implemented as a self-complete booklet. Alternatives such as laptops were considered but were not practical given the resource constraints. Self-complete booklets made the data collection process more manageable. The MCDA booklet is show in Appendix 13.

Selecting the criteria

Options are the various choices or alternatives that are considered in any decision problem. In this study the options are the six health states. Criteria are the standards by which an option can be judged to be more desirable than another (Belton and Stewart,

2001, p.1). For example, price and safety can be important criteria when purchasing a car.

The criteria in this study were based on the findings of study one (reported in Chapter 5). In study one it was found that the consequences of health states were crucial to the value of health states and that six consequences were most frequently mentioned by participants: enjoyment, relationships, independence, dignity, avoiding being a burden, and activities. These six consequences were selected as criteria in the MCDA task because they covered a range of important domains and because six criteria were practical given time and cognitive burden considerations. The six consequences were thus chosen as the six criteria.

Selecting the scoring method and anchors

The second step in designing a MCDA exercise is to develop a method to score the options. In this study the health states were scored using the global scoring method (Monat, 2009). Global scoring uses the best and worst imaginable options as anchors on a scale from zero to 100 (Monat, 2009). The participants are asked to directly state a score for each health state on a scale of 0 to 100 for each criterion. In the health economics literature this method could be described as a single attribute value function using a rating scale (Feeny et al., 2002). An alternative scale is to use 'local scoring' and use the best and worst option within the specific choice set as anchors (in this study that would be the six health states). The advantage of global scoring is that the entire range of options possible will be considered, potentially allowing for a more reflective experience. In addition local scales may cause participants to overestimate small differences between alternatives (Monat, 2009).

Labels were developed for the best and worst option imaginable anchors. It was considered most practical to use labels developed by an established and validated questionnaire, which avoids the researcher developing arbitrary labels. The six criteria

are not health domains but rather quality of life domains and therefore anchors from a quality of life questionnaire were considered most appropriate. One example of a validated quality of life questionnaire is the World Health Organisation Quality of Life assessment (WhoQOL) (The WHOQOL group, 1998; Skevington et al., 2004). The WhoQOL uses two anchors of "not at all" and "completely". The two labels were selected for the anchor points of the criteria because they covered the best and worst imaginable situations and fit with the criteria phrasing. A change was made for the lower anchor of enjoyment criterion because "not at all able to enjoy life" did not sufficiently cover the most severe states. The most severe states may include distress rather than the absence of enjoyment. An alternative to the anchor of "not at all" is a scale with the lower anchor of "completely unhappy" (OECD, 2013, p.85).

Participants were also asked to score Dead and '11111' (no problems on any domains) because that may prompt participants to consider whether a health state is better or worse than dead and how far the health state is from no health problems.

Weighing the criteria

The importance weight method (Fischer, 1995) was chosen to weigh the criteria. In this method participants select the most important criterion and give it a score of a 100, and score all other criteria relatively to the most important criterion on a scale from 0 to 100 (Fischer, 1995). This weighing method may result in bias because the weights are not explicitly based on the range of scores in each criterion (Fischer, 1995; Belton and Stewart, 2001, p.135). The swing weight method can correct this bias because it incorporates the range of the scores in each criterion explicitly (Fischer, 1995; Belton and Stewart, 2001, p.135; Ananda and Herath, 2009). However, the swing weight method takes a significant amount of time and may present participants with unrealistic scenarios (Ananda and Herath, 2009). The swing weight method was not chosen because the final aggregated MCDA score was not calculated in the group meeting and

therefore it was considered not worthwhile to spend a disproportioned amount of time on weighing the criteria. Importance weights still encourage participants to reflect on the six criteria and this helps them construct their preferences.

6.1.5 Sample size and recruitment

The sample size was calculated to be able to detect a change of 0.1 (on a scale where 1 is full health and 0 is dead) with a type I error of 0.05, type II error of 0.8, and a standard deviation of 0.3. These values were chosen based on existing literature. The value of 0.1 for detecting the change was based on the "minimal important difference" of 0.074 estimated by Walters and Brazier (2005). Walters and Brazier (2005) studied the change in EQ-5D-3L[15] value that would be considered important by patients. If that value is important to the patient, arguably it should important in comparing health state values used for cost-effectiveness analysis. A standard deviation of about 0.3 was expected based on comparable studies (Stein et al., 2006; Robinson, 2011). The type I and type II errors are based on conventional standard used in the broader scientific literature (Batterham and Atkinson, 2005). The sample size is calculated using R package 'pwr', which is based on power calculations of Cohen (1988, p.45). The required sample size with for this study was 66 participants. The aim was to recruit 70 participants and conduct 10 group meetings, for a mean of 7 participants per group. Although power could be improved with larger sample sizes, a recruitment target of 70 was similar to existing studies (e.g. (Robinson, 2011)) and was the highest number deemed feasible considering practical limitations of time, money, and recruitment sources in this PhD study.

The participants in this study were members of the general public. This is appropriate as they are used in existing valuation surveys (Brazier et al., 2005) and

15. No EQ-5D-5L value was available.

required by key reimbursement agencies such as NICE (NICE, 2013a). Recruitment was limited to Sheffield because of practical limitations. Participants were recruited by contacting the University of Sheffield staff and student volunteers email list, social groups in Sheffield, and by using the snowball method. The snowball method means that participants in the study assisted in identifying other potential participants by asking them to share details of the study with others they thought may be interested (Ritchie et al., 2003a, p.94). An email was sent out to the University of Sheffield staff and student list and members were invited to respond if they were interested in taking part in the study. Social groups and various organisations from Sheffield were also contacted. Since the sample in study one did not include many older people it was decided to recruit from two organisations in Sheffield with older members, the University of the Third Age (U3A) (Sheffield University of The Third Age, 2015) and Sheffield 50+ (Sheffield 50+, 2014). The main contact for each organisation was emailed and asked to contact their members. An advertisement was displayed on both organisations' websites. An advertisement was also placed in the monthly University of Third Age magazine (sheffieldu3a, 2015). Several other groups from the Sheffield Help Yourself (2015) website were emailed. The Sheffield Help Yourself website was the same one used in study one.

The composition of the groups was varied. The focus group literature recommends that participants have homogenous background characteristics as this facilitates discussion (Kitzinger, 1995; Spencer et al., 2003, p.190). The benefit of a homogenous group is that participants may be more willing to share their experiences, which may make the discussion more informative. Yet, this may reduce the range of different experiences and attitudes in the group. Therefore, it was felt that neither a homogenous nor a heterogeneous group composition could be said to be best. Heterogeneous group

composition was chosen because it is practically more feasible as any participant can join any group. Participants received £15 for participating in the group meeting.

6.1.6 Analysis method

The data relating to the quantitative aspect of this study was collected using the TTO booklets. The data consists of valuations of health states made using the TTO, both before and after the group meeting. The PhD student entered all data in Excel. All data analysis was conducted in R (R Core Team, 2015). All collected data by all participants were used for this study, unless the answers displayed an obvious misunderstanding of the task or the participant had left a TTO page blank. Unusable data are answers such as where the state was considered better than eight years in full health but worse than six years in full health.

To check data quality the number of inconsistencies was investigated. There are pairs of health states where a logical ordering was expected. If health state A has more problems in at least one domain and is no better in any other domain than health state B, it is logical for A to have a lower value than B. If A is not valued lower than B, the pair wise valuation is logically inconsistent (Devlin et al., 2003). Eight pairs of inconsistencies are possible between the six states in this study.

The analysis starts by calculating the values of health states using participants' TTO responses. The methods to do so are described in the next section. After this, the aggregate and individual level analyses are described.

Transforming TTO values

Participants' choices in the TTO were used to calculate the value of each health state. For states better than dead, the value for health state h_i is calculated as follows (Brazier et al., 2007, p.91):

$$V(h) = \frac{x}{t} = \frac{x}{20}$$

where V(h) is the value for the health state h, x is the number of years in full health where the respondent is indifferent between full health and t years in health state h. In this TTO t was set at 20 (see Section 6.1.3).

For the health states worse than dead the formula is (Brazier et al., 2007, p.91):

$$V(h) = \frac{-x}{t-x} = \frac{-x}{20-x}$$

where V(h) is the value for the health state h, x is the number of years in full health where the respondent is indifferent between a combination of x years full health and 20-x years in the health state h compared to being dead, and t is the combined number of years in full health and the health state h. Once again, in this TTO 't' was set at 20 (see Section 6.1.3)

The values calculated with the formula above are considered technically correct since they are derived from the participants' choices but in the literature the values for worse than dead are generally transformed. The transformation is undertaken because the untransformed negative values can be much larger than the untransformed positive values and therefore could be seen to unduly influence the mean values (Tilling et al., 2010a). In this study the lowest possible valuation is -19 but the highest possible valuation is one. The extent to which the values should be transformed has been disputed but generally all valuation studies transform their data. Two transformations are frequently used in the literature, the MVH transformation (Patrick et al., 1994) and the US transformation (Shaw et al., 2005). The MVH transformation is reported in the result section here because this has been used with the UK EQ-5D previously (Dolan, 1997a). The MVH transformation means that each year given up on the worse than dead scale has the same absolute value as a year given up on the better than dead sheet. The formula for the MVH transformation is:

$$V(h)_t = \frac{V(h)}{1-V(h)}$$

where $V(h)_t$ is the transformed value and V(h) is the untransformed value.

In the US valuation of EQ-5D the authors proposed an alternative where the distributional shape of the original untransformed worse than dead values is maintained (Shaw et al., 2005):

$$V(h)_t = \frac{V(h)}{19}$$

where $V(h)_t$ is the transformed value and V(h) is the untransformed value.

In the analyses presented in this chapter, the untransformed and the MVH transformation are reported in order to help understand the impact of the exercise on responses and to understand the implications for values likely to be used in policy making. The US transformation was undertaken but did not make a difference to the results. Therefore it was decided to report the US transformation based results in the Appendix (see Appendix 14) to reduce the amount of data presented in this chapter.

Aggregate level analysis

The first type of analysis conducted is at the aggregate level. Aggregate level analysis is required because the values individuals provide are aggregated when used in cost-effectiveness analysis (Brazier et al., 2007, p.111). The analysis undertaken at the aggregate level aims to answer two questions: (a) is there a statistically significant change in mean health state values and (b) what is the effect size of the change? One method for testing for equality of two means is the paired t-test (Field, 2009). The paired t-test is most appropriate because pre-test and post-test values are provided by the same participant and are not independent of each other. The paired t-test assumes that differences between the values are normally distributed (Field, 2009). This assumption is tested using the Shapiro-Wilk test (Field, 2009). If the assumption is not met, the paired t-test should not be used and a non-parametric equivalent has to be used. The non-parametric equivalent of the paired t-test is the Wilcoxon matched-pairs signed-rank

test, which tests the null hypothesis that the median of the differences between the pairs in the sample is zero (Field, 2009). Statistical significance is defined as a p-value of equal or less than 0.05, as per scientific consensus (Sterne and Smith, 2001; Batterham and Atkinson, 2005).

To assess the practical significance of any differences after reflection and deliberation effect sizes are computed. The effect size of a change indicates the strength of change between two values. One measure of effect size is Cohen's d, which can be calculated using the formula (Cohen, 1988, p.20):

$$d = \frac{m_2 - m_1}{SD_p}$$

where d is the effect size, m2 and m1 are the mean values for group one and group two, and SD_p is the pooled standard deviation of both groups. Cohen's d effect size can be interpreted using guidelines in the literature (Cohen, 1988, p.24). An effect size of 0.2 is considered small, 0.5 medium, and 0.8 large (Cohen, 1988, p.24). There is some disagreement about whether the guidelines are universally applicable because they only consider the size of the change and the standard deviation and do not consider the purpose of the analysis (Durlak, 2009). Nevertheless, Cohen's d provides an estimate of the size of the effect in a standardised form.

Individual level analysis

Despite the use of aggregate level values in cost-effectiveness analysis investigating individual level preferences can provide important information in investigating the effect of reflection and deliberation. At the individual level four questions are examined:

- Do the individual level values change and how reliable are the individual level values?
- Does the proportion of values better than dead (BTD)/worse than dead (WTD) change?

- Does the willingness to trade off years change?
- Does group membership, or any background variable influence likelihood of a person changing their values?

First, the number of the health state values that changed before and after discussion are described. The number of changes at the individual level and the number of changes that were more than an absolute value of (0.1) are calculated. The value of 0.1 is used based on the same reasoning as in the sample size calculation. The reliability at the individual level between the two measurements can be calculated using the intra class correlation coefficient (ICC) (Shrout and Fleiss, 1979). Several ICC measures exist, depending on whether the participants and the before and after testing time frames are assumed to be drawn from a random population or fixed, whether the scores are a mean score or of one measurement, and whether absolute or relative agreement is sought (Shrout and Fleiss, 1979). A two way mixed method ICC is calculated, as this assumes that participants and the time frame are randomly selected (Shrout and Fleiss, 1979). ICC can be calculated for agreement or consistency (Shrout and Fleiss, 1979). Agreement is determined by absolute values while consistency uses the relative values. For example, if all health values increase by 0.5, the consistency between the two measurements is high, while agreement is low. Because absolute values are important in health state valuation agreement ICC is chosen. The formula for the chosen version of the ICC is (Shrout and Fleiss, 1979):

$$ICC(2,1) = \frac{\text{BMS-EMS}}{\text{BMS+(k-1)EMS+}\frac{\text{k(JMS-EMS)}}{\text{n}}}$$

where BMS is the between-targets (i.e. participants) mean square, EMS is residual sum of squares, JMS is between-judges (i.e. two rating time periods) sum of squares, k is the number of judges, and n is the number of targets. The ICC values can be interpreted

with guidelines provided in the literature (Cicchetti, 1994). Cicchetti (1994) proposes that a value below 0.4 is poor, between 0.4 and 0.59 is fair, between 0.6 to 0.74 is good, and above 0.75 is excellent. It has to be noted that ICC may not be an ideal measure of test-retest reliability or a measure of interchangeability of two measurements (Vaz et al., 2013). ICC values are influenced by the range of scores within the group of participants and ICC values decrease when participants score similarly to each other (Vaz et al., 2013). Despite this concern, the ICC is frequently used in test-retest studies and it is used in this chapter to maintain comparability with other studies.

The proportions of participants who consider a state better or worse than dead following the discussion were compared. The proportion may be important since methods for valuing health states better or worse than dead are different (Gudex, 1994). If participants change their preferences after a reflection and deliberation this may have implications for the health state values. The proportions were compared using the McNemar test, which accounts for the paired nature of our data and is non-parametric (Field, 2009). The McNemar test tests the null hypothesis that the proportions are the same after the group meeting (Field, 2009).

Reflection and deliberation may affect a participant's general tendency towards the quality and quantity trade-off in the TTO. For example, participants may become less willing to trade off years. Studying the tendency is relevant given the adaptation literature that suggests that members of the general public may not consider the degree to which they can adapt to a health state (Brazier et al., 2007, p.116; McTaggart-Cowan et al., 2011). It's possible that reflection and deliberation would prompt participants to consider adaptation. The willingness to trade life years is measured by calculating whether all of a participant's health state values increased or decreased.

Lastly, an individual level analysis is undertaken to examine whether group membership or background variable influence likelihood of a person changing their

values. The findings of study one (Chapter 5) suggest that various conversion factors, such as age or family circumstances, were important in valuing health. For example, in study one it was noted that some participants explained how their marital status affected their preferences. It is possible, for example, that the effect of reflection and deliberation is different for single compared to married participants. It is also possible that certain group meetings produced more change than others, and it was important to investigate this.

The linear mixed effect model is appropriate for this analysis as it takes into account the clustering within the data (Snijders and Bosker, 2012, pp.43-44). Each participant provides two TTO responses, and the two responses of one participant are more likely to be similar to each other than between respondents. In addition, respondents are based in groups that may mean that values within groups are not independent. Because of these reasons, a linear mixed effect model is more appropriate than an Ordinary Least Squares (OLS) model (Snijders and Bosker, 2012, pp.43-44).

The linear mixed effect model is built in several stages. The first model fit is the 'empty model', which contains no explanatory variables but only a constant, and two random effects at the group and individual level (Snijders and Bosker, 2012, p.49). The empty model can be used to calculate an ICC value representing "the fraction of total variability that is due to the group level" (Snijders and Bosker, 2012, p.45). The higher this fraction is, the more need to explicitly model the hierarchical nature of the data. The empty model only includes a fixed intercept and two random intercepts, one for participants and one for groups (Snijders and Bosker, 2012, p.49):

$$Y_{ik} = \gamma_{00} + U_{0k} + R_{ik} + e_{ik}$$

where Y_{ik} is the value of the health state for person i in group k, γ_{00} is the constant, U_{0k} is the random variance factor reflecting the difference of the mean value of the

health state for group k, R_{ik} is the random variance factor reflecting the difference of the mean value of the health state for individual ik, and e_{ik} is the error term. The ICC can be calculated by dividing the variance of the random effects by the total variance of Y_{ik} (Snijders and Bosker, 2012, p.50).

If the empty model suggests that hierarchical nature should be modelled explicitly the explanatory variables can be added in turn (Snijders and Bosker, 2012, p.50). In model two, a variable for health state is added to model one since it is expected that health states have different mean values. Then, for model three, a covariate for time is added. Time indicates whether the value is measured before or after the group meeting. Model four includes a random slope for time. This means that different participants and groups are expected to react differently to the group meeting, for example some participants may change their values but some may not. Model five includes an interaction term between time and each health state, since different health states might change by different amounts and in different directions. Model six includes background variables for age, gender, main professional activity, marital status, children, degree, religious, own eq-5d value, and whether the participant has experienced an instance of serious illness. Finally, several interaction terms, between background variables and time, were included in model seven (for a table with the models see Table 6.11).

To test the model fit and assess whether the models improve statistically significantly with the addition of explanatory terms the log likelihood or the deviance of each model can be compared (Snijders and Bosker, 2012, p.50). The deviance test compares the difference in deviances of models using the chi-squared distribution (Snijders and Bosker, 2012, p.50). Added variables that are statistically significant are kept in the model. In order to use the deviance test models are fitted using the Maximum Likelihood (ML) method rather than the residual maximum likelihood

method (REML). The REML method only allows model comparison where the same fixed effects have been used, while ML models can be compared even if the fixed effects differ (Snijders and Bosker, 2012, p.50). The model assumes independence of the residuals, homoscedasticity, and normality of errors (Pinheiro and Bates, 2000, p.174; Snijders and Bosker, 2012, p.153). The assumptions of the model are checked using a plot of the observed values on the residuals and a qq-plot of the residuals compared to a normal distribution.

The approach above relies on pooling data of all six health states together into one model. One could argue that each health state represents a separate dependent variable and that pooling them assumes a constant effect. The assumptions of a constant effect may not be appropriate and it could be argued that therefore the six health states should be treated separately and six models should be run. To confirm whether the health state separately produce the same results to the pooled model, the same method above is applied to each health state separately.

6.2 Results

6.2.1 Sample description

A total of 62 participants took part in this study. The participants' background characteristics and the UK population figures (Office for National Statistics, 2011) are described in Table 6.2. It is not possible to estimate the response rate because the number of people who received the invitation is unknown. This is because the groups that were contacted forwarded the recruitment email and newsletter advertisement to their members and their member database sizes are unknown.

A total of 13 group meetings were held. The number of participants ranged from two to seven, with an average of 4.8 participants. The total number of participants per group is listed below in Table 6.3.

	Sample	UK population[1]
Number of participants	62	-
Number of participants with incomplete TTO valuations	5	-
Mean age	45	39
Female, n (%)	37 (60%)	50.8
Degree, n (%)	41 (66%)	27%[2]
Employed, n (%)	21 (34%)	59%
Student, n (%)	21 (34%)	8.8%
Retired (%)	17 (27%)	13%
Median EQ-5D (1st and 3rd Quartile)	1 (0.77, 1)[3]	-
Range EQ-5D	0.55 to 1	-
Has child, n (%)	21 (34%)	-
Age bracket 18 to 19	2 (3%)	3%
Age bracket 20 to 29	22 (35%)	17%
Age bracket 30 to 49	11 (18%)	36%
Age bracket 50 to 59	5 (8%)	15%
Age bracket 60 above	22 (35%)	29%

1: (Office for National Statistics, 2011), 2: Includes everyone 16 and above, 3: (van Hout et al., 2012)

Table 6.2 Background characteristics of participants and UK population

Group meeting	Participants
1	5
2	6
3	7
4	6
5	6
6	6
7	4
8	6
9	2
10	2
11	3
12	5
13	4

Table 6.3 Number of participants per group meeting

6.2.2 Descriptive analysis

All groups completed all the stages of the meeting, with each meeting lasting about two hours. Out of the 62 participants, 57 fully completed the entire process. Five participants did not complete the TTO booklets correctly but they did complete the

MCDA booklet correctly. The TTO data for these five participants were entirely removed. The errors included providing incomplete answers for a health state, leaving the page blank, or providing inconsistent answers within a health state (e.g. preferring 10 years in full health to 20 years in the health state but not 12 years in full health).

Histograms of the health states values, before and after the group meeting, using the MVH transformation are reported in Figure 6.2. The distributions do not resemble a normal distribution, exhibiting left skewness or multiple peaks. Mean values for all six health states, before and after the group meeting, are shown in Table 6.4 and plotted in Figure 6.3. As can be seen in Table 6.4, the ordering of all values is logical, with dominated health states having lower values. The mean MVH transformed values are similar to those of the cross-walk based valuation set (van Hout et al., 2012), which are reported in the last column. In the first TTO, the lowest MVH transformed value is -0.024 and the highest is 0.761. There is a large variability of values, with relatively high standard deviations ranging from 0.22 to 0.56 for before values and 0.2 to 0.58 for the after values (for the MVH transformation). The range of values and the standard deviations are similar for the first and second TTO valuations.

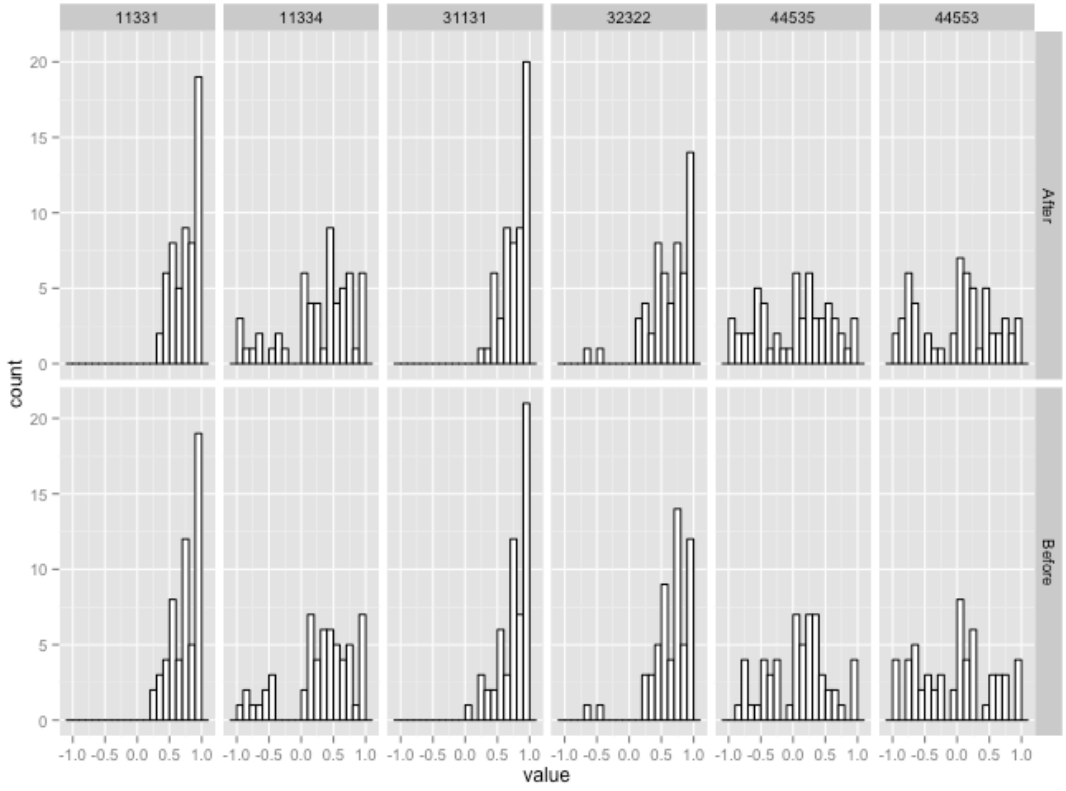


Figure 6.2 Distribution of valuations for the six health states before and after the group meeting. Valuations are MVH transformed

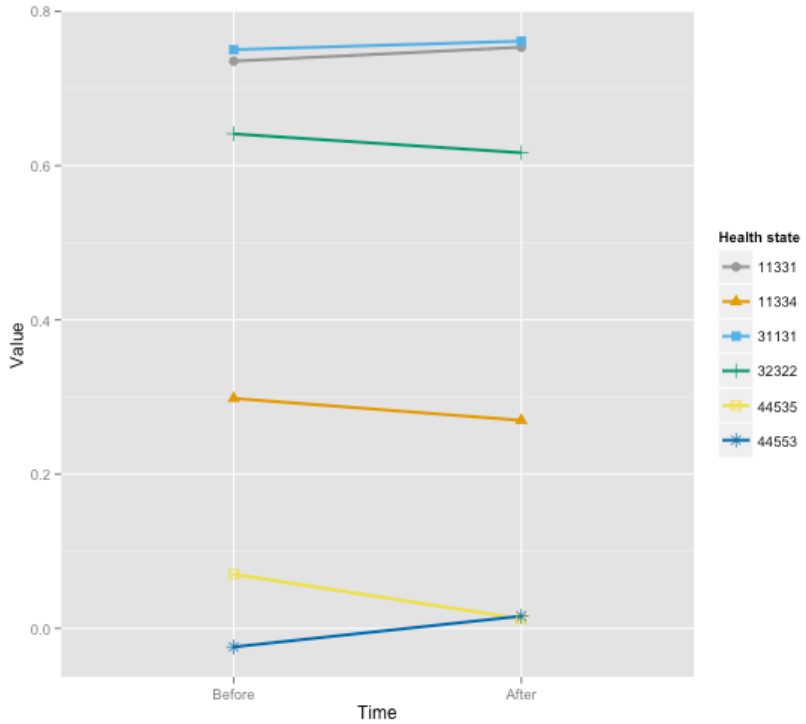


Figure 6.3 Line chart of mean health state values before and after group meeting for MVH transformed values

Health state	MVH before Value (SD)	MVH after value (SD)	Change (SD)	Before value (SD)	After value (SD)	Change (SD)	Cross-walk value[1]
44535	0.07 (0.47)	0.012 (0.56)	-0.058 (0.08)	-0.289 (1.35)	-1.243 (4.04)	-0.954 (2.69)	-0.020
11331	0.736 (0.22)	0.754 (0.2)	0.018 (-0.01)	0.736 (0.22)	0.754 (0.2)	0.018 (-0.01)	0.727
32322	0.641 (0.32)	0.617 (0.35)	-0.024 (0.03)	0.609 (0.47)	0.58 (0.51)	-0.029 (0.04)	0.573
31131	0.75 (0.24)	0.761 (0.2)	0.011 (-0.04)	0.75 (0.24)	0.761 (0.2)	0.011 (-0.04)	0.76
11334	0.298 (0.52)	0.27 (0.54)	-0.028 (0.02)	-0.316 (2.89)	-0.732 (4.01)	-0.416 (1.12)	0.476
44553	-0.024 (0.56)	0.016 (0.58)	0.04 (0.01)	-1.392 (4.21)	-1.256 (3.86)	0.136 (-0.34)	-0.118

1: Based on (van Hout et al., 2012)

Table 6.4 Mean health state values before and after group meeting, difference between the two, and cross-walk values. Both MVH transformation and untransformed values are reported

Table 6.4 shows a large difference between the untransformed and transformed values for the three health states with the lowest values (health state 44535, 11344, and 44553), but small or no differences for the other three health states. For example, the untransformed and MVH transformed values for health state 32322 are 0.61 and 0.64, while for health state 44553 they are -1.39 and -0.02. The difference occurs because the transformation only affects negative values. The number of negative values was 78% for state 44553, 50% for state 44535, 21% for state 11344, 4% for state 32322, 0% for state 31131, and 0% for state 11131. Health states that were valued worse than dead (hence have negative values) relatively often are more affected by the transformation.

Table 6.5 displays the number of times inconsistencies occur. The pairing of health state 44535 and 11344 has many more inconsistencies than others. This was noted after the first five group meetings and the order of 44535 and another arbitrary chosen state was reversed (32322). The states were reversed to see whether the order of the health states could be a cause of the inconsistency. In the last six group meetings the inconsistencies are reduced as shown in Table 6.5. The number of inconsistencies for the other seven pairings is low with just one or two cases.

Inconsistency	Percentage of all valuations	Before switch	After switch
44535 higher than 11331	0%	0%	0%
44535 higher than 32322	2%	3%	0%
44535 higher than 31131	2%	0%	4%
44535 higher than 11334	21%	26%	13%
11334 higher than 11331	2%	3%	0%
44553 higher than 11331	2%	0%	4%
44553 higher than 32322	2%	0%	4%
44553 higher than 31131	2%	0%	4%

Table 6.5 Percentage of inconsistent valuations

6.2.3 Aggregate analysis

The first research question for this chapter was to assess whether mean health state values changed before and after reflection and deliberation. Table 6.4 shows the changes in health state values. For untransformed values, changes before and after discussion range from -0.954 (health state 44535) to 0.136 (health state 44553), with a mean absolute value of 0.26. After applying the MVH transformation, they range from -0.058 to 0.04 with a mean absolute value of 0.03. The states ranked by mean MVH transformed health state values before the discussion as: health state 31131, 11331, 32322, 11334, 44535, and lastly 44553. After the group discussion states the ranking of 44535 and 44553 reversed.

Distributions of the changes for each health state are shown in Figure 6.4. The Shapiro-Wilk test indicates that the differences for all of the health states cannot be assumed to be normal (all p values <0.01). As the difference between the pre and post values cannot be assumed to be normal, the non-parametric Wilcoxon test has to be used.

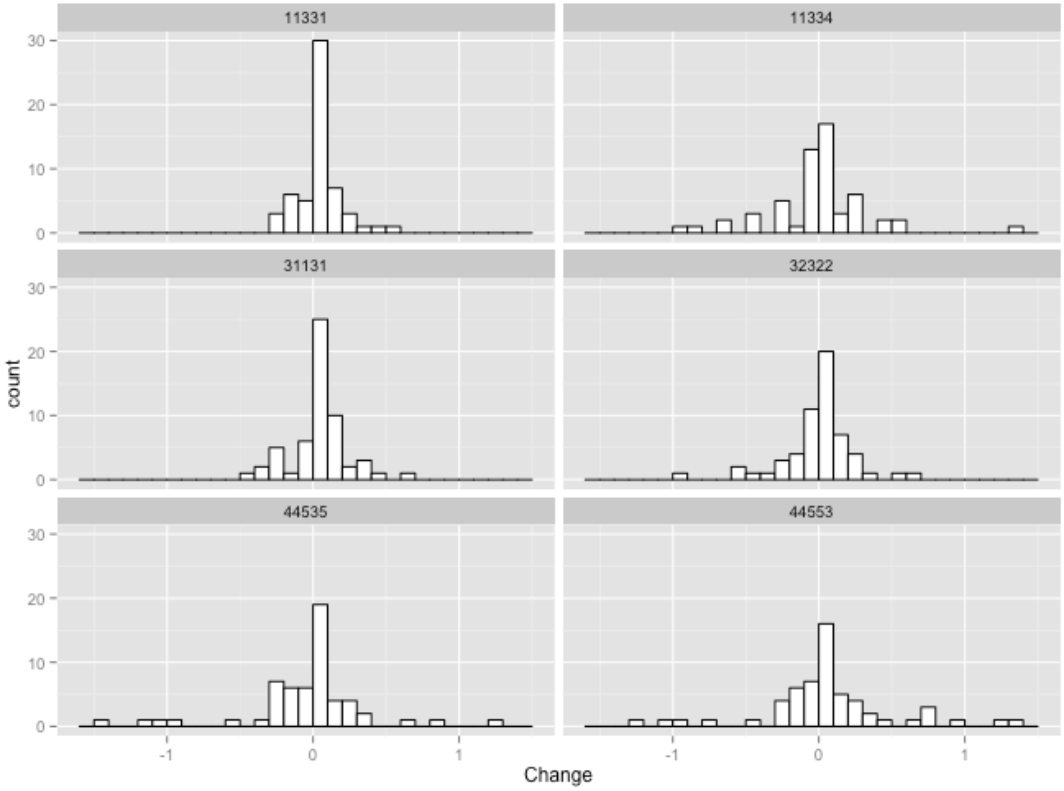


Figure 6.4 Distribution of changes for six health states, using MVH transformation

Results of the Wilcoxon test are displayed in Table 6.6. The null hypothesis that the median differences between the health state values before and after discussion is zero cannot be rejected for any of the six health states because all p-values are above the threshold. There are no statistically significant changes for any health state values[16].

The effect sizes of the changes for the health states are reported in Table 6.7. These range from 0.05 to 0.11 for the MVH transformation. According to the guidelines of Cohen (Cohen, 1988, p.24) these effect sizes are considered below 'small' as they are below 0.2.

[16] Similar to the Wilcoxon test, the results from the parametric paired t-test do not indicate a statistically significant difference (p-values for MVH transformation range from 0.30 to 0.67).

Health state	Untransformed p-value	MVH p-value
44535	0.61	0.27
31131	0.63	0.63
32322	0.49	0.63
11331	0.52	0.52
11334	0.43	0.49
44553	0.88	0.58

Table 6.6 P-value for null hypothesis of no change after group meeting, calculated using Wilcoxon test

Health state	Untransformed Cohen's d	MVH Cohen's d
44535	0.32	0.11
31131	0.09	0.09
32322	0.06	0.07
11331	0.05	0.05
11334	0.12	0.05
44553	0.03	0.07

Table 6.7 Effect size of changes in health state values, calculated using Cohen's d.

The aggregate level analysis suggests that there is no statistically significant difference between the health state values before and after the group meeting and that the changes are of lower than small effect size. The next section reports the results of the individual level analysis.

6.2.4 Individual level analysis

Most participants changed their health state values after the group meeting, but this was not done in a systematic fashion. The number of changes at the individual level is reported in Table 6.8. The first column of Table 6.8 reports the number of changes made by participants for each health state, while the last column displays the number of changes that were more than an absolute value of 0.1.

On average, 75% of participants changed their values for a health state and 72% of the changes were greater than an absolute value of 0.1. The third and fourth columns of Table 6.8 display the number of participants who increased or decreased their valuation post group meeting. For most states the two numbers are similar, a total of 130 changes

were increases in valuation and 125 were decreases. The large number of changes at the individual level does not appear systematic.

Health state	Changes values (% of all valuations)	Changes bigger or equal than absolute value of 0.1 (% of all changes)	Increased valuations	Decreased valuations
44535	47 (82%)	34 (72%)	22	25
31131	35 (61%)	24 (69%)	21	14
32322	42 (74%)	29 (69%)	19	23
11331	39 (68%)	29 (74%)	24	15
11334	47 (82%)	32 (68%)	21	26
44553	45 (79%)	35 (78%)	23	22

Table 6.8 Number of total changes, changes greater than absolute value of 0.1, count of increased and count of decreased valuations per health state. Values are MVH transformed

Reliability of individual level values can be assessed using ICC values. Table 6.9 displays the ICC values for each health state. These ICC values display a measure of reliability between the two valuations. The ICC for MVH transformation ranges from 0.62 to 0.79. Compared to guidelines provided by (Cicchetti, 1994) the ICC values are considered in the range of 'good' to 'excellent'.

Health states	Untransformed values ICC	MVH transformed values ICC
44535	0.37	0.68
31131	0.77	0.77
32322	0.81	0.73
11331	0.62	0.62
11334	0.83	0.79
44553	0.51	0.68

Table 6.9 Intraclass Correlation Coefficient statistics for six health states

The pattern of better or worse than dead valuations is similar before and after the group meeting. Table 6.10 displays whether participants who valued a state better or worse than dead also did so after the group meeting. Participants who valued a health states BTD before the group meeting tended to value it BTD after the group meeting, and the same applies to WTD values. For example in health state 44535, 86% of participants gave the same BTD or WTD valuation. The proportions of better or worse

than being dead can be formally tested with the McNemar test. The hypothesis that the proportion of states BTD or WTD is the equal after group discussion cannot be rejected. There is no statistically significant change in the proportion of individuals valuing a state better and worse than dead.

Health state	Percentage of participants not changing from better than dead or worse than dead valuation[1]	p-value
44535	86%	0.29
31131	100%	N/A
32322	96%	1.00
11331	100%	N/A
11334	91%	1.00
44553	82%	0.34

1: Health states valued at zero are included in better than dead

Table 6.10 Percentage of participants who maintained better than dead or worse than dead valuation after group meeting and p-value of no change calculated with McNemar's test

For 18 of 57 (32%) participants changes in valuation were all only in a positive or all only in a negative direction. The number of participants with all positive changes is close to the number with all negative changes. For 10 participants all their changes were negative, while for eight participants all their changes were positive. Reflection and deliberation does not appear to change willingness to trade between the quality of life and quantify of life for the participants as a whole.

The individual level results indicate little systematic change in the health state values, but to look at the data more rigorously the data can be modelled using the linear mixed effects model.

Linear mixed effects model analysis

The empty model shows a high level of variance due to groups and individuals. The random effect part of the empty model shows the variance that is due to group membership (variance of 0.01), individual level (variance of 0.06), and the residual (variance of 0.21). The ICC calculated from the empty regression suggests that 25% of

the variance is attributable to differences between groups and subjects, with subjects having a larger variance than groups. The level of variance due to group membership and individuals indicates that a multi-level model is indeed preferable to a model that ignores the hierarchical nature of the data.

Time and background variables are not found to be significant predictors of health state values. The structure of the models estimated and their deviance test results are displayed in Table 6.11. Column two and three display the fixed and random part of the model. The first model contains no background variables, only a random intercept for groups and individuals. The addition of a term for each health state, in model two, statistically significantly improved the first model (reducing deviance by 478 with a difference of 5 degrees of freedom, resulting in a p value of <0.01). However, none of the other additions statistically significantly improve the model.[17] The additions of time and background variables in model three, four, and five only marginally reduced deviance and were not statistically significant (p-value of 0.77, 0.34, and 0.85). This test is in agreement with the Wilcoxon t-test that values after reflection and deliberation are not statically significantly different.

17. Model 7 was a combination of various interaction terms between all the variables. This model description was not shown as no statistically significant improvements in fit were made.

Model	Fixed effects	Random effects	Deviance difference
1	-	Group and individual	-
2	Health state	Group and individual	478.19[1]
3	Health state Time	Group and individual	0.09
4	Health state Time	Time (as a random slope) Group and individual	4.54
5	Health state Time Interaction of health state and time	Group and individual	1.98
6	Health state Time Interaction of health state and time Age Gender Main activity Marital Children Education Religious Own EQ-5D[2] Instance of illness	Group and individual	2.84

1: P-value <0.05, 2: Based on (van Hout et al., 2012)

Table 6.11 Model fit comparison

For information, the coefficients of model 5 (time, health state, and their interaction as fixed effect and a random intercept for groups and individuals) are shown in Table 6.12. The five dummies for the health states indicate the difference in value for the five health states compared to heath state 44535 at time zero. The time variable indicates the change in value for health state 44535 after group discussion, while the interaction terms indicate the change for the other five health states compared to the change in health state 44535.

The coefficient for time has a small t-value, as do their interaction terms. The t-values for the health states are much larger. The intercept term denotes the mean value of health state 44535 at time zero (before the reflection and deliberation). The estimates for the time terms are all relatively small, ranging from -0.06 to 0.10.

Random effect			
	Name	Variance	Std.Dev.
Individual	(Intercept)	0.07	0.26
Groups	(Intercept)	0.01	0.11
Residual		0.10	0.31
Fixed effects			
	Estimate	Std.Error	t value
(Intercept)	0.08	0.06	1.22
Time (i.e. after group valuation)	-0.06	0.06	-1.00
31131	0.67	0.06	11.49
32322	0.57	0.06	9.86
11331	0.68	0.06	11.75
11334	0.23	0.06	3.94
44553	-0.09	0.06	-1.63
After*31131	0.08	0.08	0.93
After*32322	0.03	0.08	0.41
After*11331	0.07	0.08	0.84
After*11334	0.03	0.08	0.36
After*44553	0.10	0.08	1.20

Table 6.12 Coefficients of model 5

Model diagnostics for model 5 are reported in Figure 6.5 and Figure 6.6. A visual inspection of the level one residuals shows that there may be problems with heteroscedacity because the points are not randomly scattered with a constant spread. The QQplot indicates that errors may be non-normally distributed. Inferences from this model need to consider this.

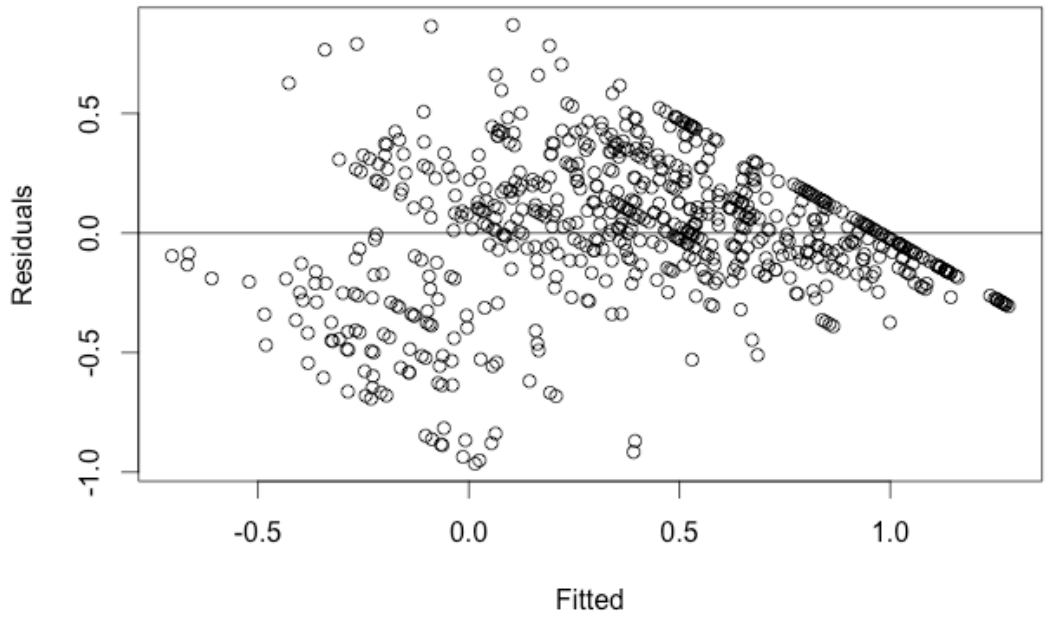


Figure 6.5 Scatter plot of the level 1 residuals versus the fitted values for model 5

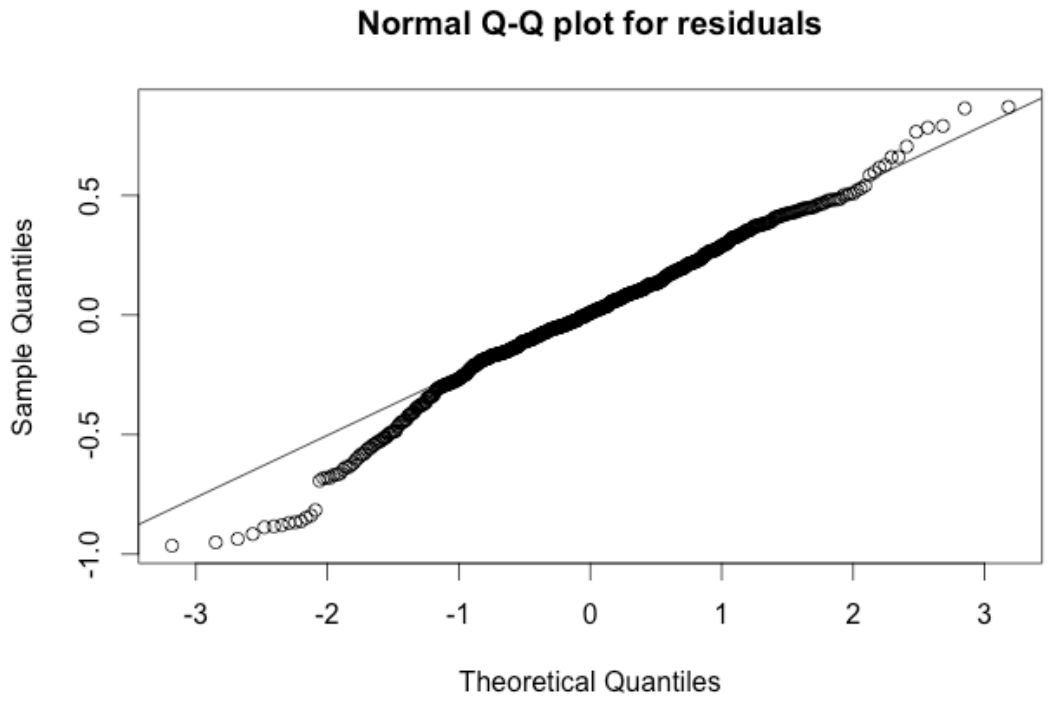


Figure 6.6 Normal plot of the level 1 residuals for model 5

The same modelling approach was used for the six health states separately. The findings are in accordance with the pooled model and no statistically significant effect

of time or any other background variable was found. For information, the results of the models are included in Appendix 15 and Appendix 16.

The results of the linear mixed model indicate that the addition of time or background variables did not improve the model and was not able to predict the health state values of the participants. The finding that the time variable was not statistically significant from zero is in accordance with the result of the Wilcoxon signed-rank test. Similar to the Wilcoxon signed-rank test, the mixed effect model suggests that the effect of the group meeting on health state values is not statistically significant. In addition, there are no covariates that explain data at a statistically significant level.

6.3 Discussion

6.3.1 Summary

The findings of this study indicate that mean level health state values do not statistically significantly change after a structured reflection and deliberation exercise. The largest change using the MVH transformation was -0.06, but compared with a standard deviation of 0.52 it represents a very small effect size. The linear mixed effect analysis confirmed the finding because time was not significant predictors of valuation. No background variables were statistically significant in predicting health state values.

There are, however, relatively large and frequent changes to the preferences of participants at the individual level. The changes appear to be non-systematic and tend to cancel each other out. The number of participants making only positive changes was the same as the number of participants making negative changes, which does not indicate that adaptation or the general willingness to trade off life years has changed. The proportion of values being BTD or WTD remained similar in the first and second TTO for all six health states.

6.3.2 Comparison to test-retest studies

The extent of the changes in health state values in this study can be better understood by comparing them to test-retest studies of the TTO. Test-retest studies re-administer the TTO after a period of time (for example two weeks) without any intervention and therefore indicate the degree of randomness in the TTO (Brazier et al., 1999b). They assume that no significant changes have taken place in between the valuations. Three such studies were cited in Brazier et al. (1999b), which reviewed the reliability and validity of various health status measures. From the studies reported by Brazier et al. (1999b) those using the EQ-5D, valued with the TTO, and who reported an ICC value were selected. Two further studies were identified by searching Scopus for "test-retest" studies using "TTO" published after the year 2000. Table 6.13 displays the reliability as measured by ICC values found in this study and those of the test-retest studies.

The ICC values of this study are similar to test-retest values found by Robinson (2011) and slightly lower than others. The ICC values indicate that the group meeting causes as much, or slightly more, change as simply re-testing participants after two or more weeks. There was only at most 100 minutes between the two valuations in this study and therefore it may be surprising that the reliability values are lower than test-retest studies with longer time intervals between administration. This would suggest that the reflection and deliberation does prompt changes in participant's preferences at the individual level, even though it appears to be largely random.

Author	Time between tests	States	Sample size	Instrument	ICC value
Badia et al. (1999)	One to four weeks	13	50	Eq-5D-3l, TTO with props	0.84
Dolan et al. (1996b)	10 weeks	43	221	EQ-5D-3L, TTO with props	0.73
Gage et al. (1996)	Median of 6.9 weeks	3	23	Stroke description, TTO method not clear	0.67 to 0.92
Robinson (2011)	Two to four weeks	12	171	EQ-5D-3L self completion	0.4 to 0.88 (median of 0.66)
Buitinga et al. (2011)	Two weeks	1	29	Computer TTO using participants' own health state	0.88
This study	Less than two hours	6	57	EQ-5D-5L self completion	0.6 to 0.78

Table 6.13 Intraclass Correlation Coefficient values from the literature compared to findings from this study

6.3.3 Comparison to literature

The results of this study provide further evidence to the one study in the literature (see Chapter 3) that a group-based reflection and deliberation exercise does not statistically significantly change health state values (Stein et al., 2006). In the studies reviewed in Chapter 3 it was reported that four group-based studies (Krabbe et al., 1996; Akunne et al., 2006; McIntosh et al., 2007; Robinson and Bryan, 2013) showed statistically significant changes of health state values. Yet it was argued that perhaps the results of those studies are not generalizable because of different aggregation methods, valuation methods, study designs, and classification systems. Indeed the findings of this chapter would suggest that this may be the case.

The voting method used by Krabbe et al. (1996) is unconventional and is perhaps better compared to consensus-based methods. It is hard to judge how their method of interactive voting would influence mean-based aggregation reflection and deliberation.

Similarly, Akunne et al. (2006) reported that consensus based group valuations were higher (health states were seen as more severe) than individual values but the change in values may have been due to requiring consensus by the group. McIntosh et al. (2007) found that for five of twelve states, there was a statistically significant impact of the discussion. Two possible explanations are the study design and health state classification system. The participants were given a chance to change their values rather than asked to re-value the states. This resulted in fewer individual level changes than in this study and on average only 16% of health state values were changed. Therefore, the standard errors of the mean of the differences were low and combined with the fairly large sample size of 139 can lead to a statistically significant change. Robinson (2011) reported significant changes after a group discussion but the valuation method used was the person trade off method (PTO). The PTO has lower test-retest reliability than the TTO (Robinson, 2011) and involves equity considerations not present in the TTO (Damschroder et al., 2005a) and it is possible that the changes are caused by the use of PTO rather than the nature of the group exercise.

The findings here support the studies that show that mean values are not likely to change. It would suggest that the perhaps the results from the other studies are not generalizable to the TTO, EQ-5D, and mean-based aggregate valuations.

6.3.4 Implications for practice

Chapter 2 presented arguments why reflection and deliberation may be needed to guide participants in constructing their health state preferences. The results here would suggest that even if a group meeting encourages reflection and deliberation it does not change aggregate health state values. The high number of changes at the individual level may indeed suggest preference construction at the individual level but ultimately it does not make a practical difference for aggregate values used in cost-effectiveness analysis.

Although no clear yardstick is available to judge when health state values changes are practically significant, the small effect size would suggest that deliberative method would not differ much from existing methods. The assumptions of complete preferences underlying choice-based methods may be violated but is not consequential at the aggregate level. The use of reflection and deliberation does not improve on conventional methods for producing mean health state values and the assumption of completeness is not shown to be invalidated at the aggregate level.

For use in patient decision-making, however, the results suggest that the individual preferences may not be reliable enough. As in cost-effectiveness, there is no established measure of how precise individual preferences need to be to be valid in patient decision-making. However, if utility differences of 0.1 are considered important, the number of changes of this magnitude in this study suggests that participants' health state preferences are too unstable to be used for patient decision-making.

6.3.5 Limitations of study

One limitation of this study is the recruiting method. A probability based recruiting method would have been better as a more representative sample could have been recruited (Bryman, 2012, p.191). In addition, group meetings held during weekends may have helped to recruit the working population between 30 to 59, as their availability to participate in a two hour session during the day is likely to be limited. For this PhD this was not practical due to time and financial constraints. Compared to the UK population (Office for National Statistics, 2011) the sample contains more females, those with degrees, students, and retirees, but fewer employed people. Compared to the UK population (Office for National Statistics, 2011) those aged 60 and above were oversampled, while 30 to 59 year olds were under sampled. It's difficult to assess how the sample affects the results. It's possible that younger participants have less experience

with ill health and their preferences may be less stable than older participants. On the other hand, instability may be due to cognitive overburden and perhaps those enrolled in university and those with degrees may be less overburdened than the average member of the general public. Although the analysis did not reveal any background variables to be significant in explaining changes it is possible that a representative sample would reveal different results. Most studies cited in the literature do not use representative samples, as this research area remains an exploratory area of research. Overall, care should be taken to generalise these findings to a larger population.

The sample size in this study is a limitation. There is more variation in size of participants per group than planned because there were last-minute cancellations by participants. It was envisaged that at least seven members per group would be present but most groups were smaller. Although the discussion was not shorter in the smaller groups and the quantitative evidence (including the mixed effect model) did not reveal an association between group size and changes to health state values, it may be that smaller groups contained less diverse views and therefore resulted in fewer reasons for participants to change their values. Larger sample sizes would also have increased the chance of distinguished changes at a statistically significant level. For example, the sample size required to find the change of -0.058 for health state 44535 statistically significant (given the standard deviation found, assuming a power level of 0.8, and an alpha of 0.05) is 351 participants. Practical considerations limited the sample size that was possible. However, even if statistically significant results were found with a larger sample size, it is not clear that they would be practically significant, given their small absolute size. Nevertheless, a larger sample would have provided more conclusive evidence.

The pretest-posttest design could be improved. There is the possibility to include a control group and account for the possibility that changes are due to random error,

where this study only used comparisons to re-test studies. Another limitation of this design is that the act of pretesting could affect the effect of the intervention. One solution is to use a control group with no pre-testing or a more sophisticated design such as the Solomon four-group design (Dimitrov and Rumrill, 2003). The Solomon four-group design uses four groups, essentially combining a pretest-posttest with control with a posttest with control design (Dimitrov and Rumrill, 2003). This design allows a researcher to distinguish the effect of the intervention and the interaction between the pre-testing and the intervention (Dimitrov and Rumrill, 2003). However, practical resource constraints did not allow for the use of control groups.

The collected health state valuation data contained a number of inconsistencies. The pair of health state 44535 and 11334 had more inconsistencies than the others. Two possible reasons are the levels of the 'anxiety and depression' domain and ordering effects. Health state 44535 contained 'extreme' while health state 11334 contained 'severe' anxiety. Valuations of the two levels in this domain have previously been shown to be inconsistent (Devlin et al., 2014). There may thus be particular problem for participants in distinguishing between severe and extreme anxiety or depression. Another possibility is that the inconsistencies are due to an order effect and it is possible that health state 11334 received lower scores than it would have done otherwise, because it came after a mild state. Ideally, the order of each state should have been randomised but this was not feasible using the pen and paper method because each health state was printed as part of the entire booklet. The position of state 44535 and state 32322 in the TTO booklet was reversed for the last seven group meetings. The order change reduced the level of inconsistencies, which would suggest that ordering effect is at least part of the reason. Reversing the health states is not ideal and ideally health states should have been randomised for each booklet. However, overall the data is of adequate quality, for example the health state values are similar to cross-

walk based 5L valuations and the mean values of the health states are logically consistent. In addition, the inconsistencies are not material to the aims of this study because differences between valuations within health states are not affected by inconsistencies.

A self-completion TTO with top down titration was used in this study. There is increasing recognition that the design of the TTO will influence responses and thus health state values (Attema et al., 2013; Versteegh et al., 2013). Therefore the TTO design in this study could influence the results, but it is not definitely possible to say how. Top down titration methods may produce higher values than other iterative methods (Lenert et al., 1998; Attema et al., 2013). If this is occurs, then aggregate values of the second TTO would be more similar to first TTO values if participants continued to use a smaller part of the scale because of top down titration than if another iterative method was chosen. In addition, there is some indication that self-completion methods without props may produce less reliable values than those with props (Dolan et al., 1996b). This would result in more change in the second TTO compared to the first TTO than if TTO with props were used. Thus, the overall effect of using this design of TTO is difficult to judge.

The linear mixed effects model used in this study, similar to that used by Damschroder et al. (2008), overcomes some of the problems of an OLS model but still contains some assumptions. In particular this model could be improved by taking account of the heteroscedacity and the bounded nature of the valuation data. Although the adjustment is not likely to change the inference that deliberation had no effect, the size of the coefficients would be more reliable if the assumption of the model were more fully met. Although this model is well-established in the health economics literature (Brazier et al., 2007, pp.146-147), there is scope for further sophisticated models such as generalised linear mixed models.

The MCDA task in this chapter cannot be considered choice-based according to the definition previously given (i.e it is not TTO, SG, or DCE). It was emphasised that the superiority of choice-based methods are based on both practical and theoretical reasons (Torrance et al., 2001; Brazier and McCabe, 2007). The practical issues with the VAS are also present in the MCDA because the methods used for MCDA in this thesis, direct scoring of the options and importance weight, resemble the VAS more than the TTO or SG. The rating scales used to score each health state on a criterion can be considered an individual VAS exercise and the importance weights method directly ask respondents to provide a score rather than trade off anything. The swing weight method is designed to reflect opportunity cost and to encourage respondents to explicitly trade off improvements in one dimension with improvements in another dimension. Even the swing weight method still differs from choice-based tasks because respondents are asked for a direct estimate of an equivalent value and the values are not inferred from a choice-based task (Belton and Stewart, 2001, p.135), although the swing weight procedure could be adjusted to guide participants to a point of indifference. What is likely to more important if the actual MCDA values are to be used is the practical validity of the values elicited using the swing weight method. However, the disconnect between MCDA and choice-based methods is less of a concern in this thesis because MCDA is here only used as a tool for reflection and deliberation and not for valuation, and the use of VAS as a prop is accepted in the literature (Torrance et al., 2001; Brazier and McCabe, 2007).

6.3.6 Future research

Future research could first focus on understanding the reasons why participants may change their preferences after reflection and deliberation but do not do so systematically. Better understanding this can highlight why the studies show different

findings. It could also help interpret the need for preference construction and highlight which aspects of reflection or deliberation are most useful to participants. This is the focus of the next chapter.

6.4 Conclusion

This chapter reported on a study testing the effect of a structured reflection and deliberation group-based exercise on health state values. This study found that participants frequently change their preferences at the individual level but that the aggregate level health state values did not statistically significantly change. The changes at the aggregate level that took place were small according to published standards of effect size. The method of reflection and deliberation examined here does not appear to produce different valuations to those of conventional TTO valuations that do not incorporate reflection and deliberation. The findings suggest that the assumption of completeness cannot be rejected. The use of conventional methods with one-off interviews, for purposes that use aggregated valuations, was not invalidated by this study.

The question remains why the health state values changed so much at the individual level. To better understand the results of this study it is helpful to understand the nature and content of the group discussion. This can be accomplished by a mixed methods study, which uses the audio recordings of the group meetings to explain the quantitative results from this chapter. The mixed methods study is the subject of the next chapter.

Chapter 7

Explaining the effect of reflection and deliberation on health state values

The objective of study two was to determine whether there is a need for reflection and deliberation in health state valuation. Part of this objective is to study the quantitative effects of reflection and deliberation. This was undertaken in Chapter 6. But an estimate of the quantitative effects alone cannot determine the role of reflection and deliberation in health state valuation. It is also informative to understand the reasons for the quantitative effect of reflection and deliberation on health state preferences.

In Chapter 6 a group meeting was described where members of the public valued health states before and after a reflection and deliberation task. The reflection task consisted of completing a MCDA booklet, which encouraged participants to reflect on living with different states of ill health. The deliberation task was a group discussion with other members of the public on the MCDA booklet. It was found that there were no statistically significant changes to mean health state values after the reflection and deliberation task.

Different reasons could explain the non-statistically significant changes in the quantitative aspect of study two and the implications of the findings may depend on the reasons. For example, if the deliberation did not reach saturation or if important aspects of participants' preferences were not discussed then a different design of reflection and deliberation is required; if participants find reflection and deliberation difficult then reflection and deliberation is not helpful in constructing preferences. A deeper understanding of the group meeting will help provide more considered assessment of the role of reflection and deliberation.

This chapter attempts a mixed methods study to explain the reasons for the quantitative findings of study two (reported in Chapter 6). A mixed methods study is a study where both quantitative and qualitative findings are combined (Creswell, 2014, pp.14-15) and is appropriate when both qualitative and quantitative data are needed to answer a research question (O'Cathain et al., 2010). The qualitative data in this chapter are the recordings of the discussion period of each group meeting and the quantitative data are the TTO valuations of the participants. In this chapter the qualitative data and quantitative data are integrated with the aim to explain the quantitative findings. Using both a qualitative and quantitative component together does present challenges and there is little existing mixed method research of this kind. Therefore a particular novel aspect of this chapter is the application of methods to combine qualitative and quantitative data.

The next section describes the methods used in this chapter to integrate quantitative and qualitative data.

7.1 Methods

7.1.1 Mixed methods design

Two of the studies cited in the literature review (see Chapter 3) on deliberation and reflection have included a formal qualitative analysis (Stein et al., 2006; Robinson and Bryan, 2013). Both interviewed participants after the group meetings to obtain the participants' views on the meeting. This means that participants have to recall the group discussion but also provide their interpretation of the group meeting. To explain what happened during the group meeting in study two (reported in Chapter 6), it may be useful to study it directly and in this chapter recordings of the deliberation period are used.

The mixed method study in this chapter can be classified as a concurrent mixed design, which means the qualitative and quantitative data were collected at the same time (Creswell, 2014, pp.231-232). The study is concurrent because, as indicated in the last paragraph, it was desirable to collect data on the deliberation period, which means that the qualitative data is collected at the same time as the quantitative data was collected. The triangulation protocol is used to integrate the qualitative and quantitative component (O'Cathain et al., 2010). In the triangulation protocol both qualitative and quantitative data are analysed separately and two sets of findings are developed. The two findings are then compared. Triangulating requires researchers to compare the findings from each component of a study and "consider where findings from each method agree (convergence), offer complementary information on the same issue (complementarity), or appear to contradict each other (discrepancy)" (O'Cathain et al., 2010, p.1147). By comparing the findings the researcher can come to an overall conclusion on the relevant research question.

A diagram of the study design is provided in Figure 7.1. The steps are described in the next five sections.

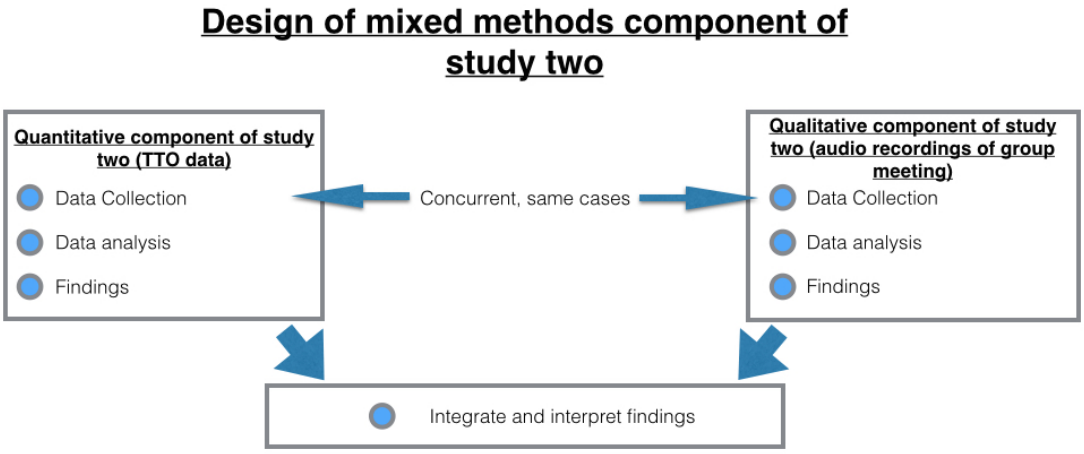


Figure 7.1 Design of mixed methods component of study two

7.1.2 Analysis method

The data in this mixed method study consist of two components: audio recordings of the group meetings and the two rounds of TTO data completed by the participants.

Five steps were followed to implement the triangulation protocol and to combine the two data components. First, a qualitative analysis of the recordings of the group meetings was conducted using Framework. Second, the themes that may explain the changes in participants' health state values were selected. Those selected themes were then summarised and a quantitative hypothesis for each of those summaries was proposed. Finally, evidence was sought for that hypothesis in the quantitative data.

Stage 1: analysis of the qualitative data

Themes were developed to describe the qualitative data. Framework (Ritchie et al., 2003b) was used for qualitative data analysis, as was done in study one (see Chapter 5). The PhD student transcribed each audio recording verbatim. The transcripts were reviewed and each significant idea was coded. These codes were organised into themes and the themes were applied to all transcripts, which is called 'indexing' the transcripts (Spencer et al., 2003). This process was iterative, as after each transcript the codes were revised and reorganised. All transcripts were reviewed three times.

Stage 2: selection of relevant themes

The aim of the analysis was to explain the changes in participants' health state valuations in the second TTO exercise and thus only the themes that could explain these changes were further investigated. For example, comments indicating that participants were surprised might explain participants changing their values, but off-topic comments were not likely to explain changes in participants' health state values and thus were not further investigated.

Stage 3: summarising the themes

Each theme was summarised, providing a short description of the ideas that were subsumed under that theme. For most themes narrative summaries are presented but the theme of 'participants view on the health states' was summarised by quantifying all the qualitative data.

The aim of the theme was to describe how the first TTO valuations of the states compared to the deliberation about each state, in order to establish whether some health states were overvalued or undervalued (e.g. whether a state was discussed very negatively but valued highly). Three steps were followed:

1. First, each mention of a health state was classified as positive or negative. A positive statement does not mean that the ill health state was desired but rather that the statement described the health state in a relatively positive way, for example describing possibilities of adaptation or coping.
2. Second, to obtain a ranking based on deliberation, the total number of negative mentions was subtracted from positive mentions and this number was ranked from highest to lowest. Higher numbers indicate that the state was discussed more positively.

The themes that emerged from the qualitative analysis could be summarised at three levels: individual, group, or aggregate. In the individual case, each individual's comments can be analysed separately. Individual level analysis may not be appropriate because an individual's valuations may reflect both what they said and what they heard from other members in the group. Alternatively, each group discussion can be analysed separately. Group level analysis is difficult due to practical constraints because of the large number of groups and is arguably less necessary if the findings from the groups are similar. If the findings are similar, it is possible to aggregate all the groups and in effect make the assumption that everyone was in one group. Therefore, the appropriate level of analysis

depends on the similarity of findings amongst the groups. The choice is justified in the results section of the chapter.

Stage 4: developing quantitative hypotheses

Each summary of each theme was reviewed and a hypothesis was developed about what that theme would suggest about changes to preferences of the participants. The hypotheses connect the qualitative and quantitative components by providing quantitative predictions based on qualitative findings. Testing the hypotheses allows the two components to be integrated.

Stage 5: assessing evidence for hypotheses

Each hypothesis was compared to the quantitative TTO data. In this way the convergence, complementarity, or discrepancy between the two components was assessed and the qualitative findings which may explain the quantitative findings can be explored (O'Cathain et al., 2010).[18]

7.2 Results

7.2.1 Findings of qualitative data

Eight themes were found after analysing the qualitative data. The themes are shown in Table 7.1, along with a definition and a quote for each theme. A detailed discussion of these themes follows later in the chapter. Note that in all quotes in this chapter, the words 'ED', 'YM', 'IR', 'YC', 'GY', 'AU', 'NA', and 'UI' refer to arbitrary labels with no meaning. They were used so that participants could more easily refer to a specific state during discussion than to read out the entire health state (see Table 6.1 on page 129).

18. One important limitation is that the convergence of the two components is not definite evidence that the qualitative explanation is correct because some third explanation of the quantitative findings may be the correct one. This is further discussed in the limitation section.

Themes	**Definition**	**Example quotation**
Agreement or disagreement	Where participants indicated agreement or disagreement with a previous expressed thought	"See I disagree with that…but possibility because, I uhm, because I live alone, I've got no family around, but your friends will stick by you."
Grouping of health states	Where participants commented on similarity between heath states	"Well I had a clear divide, I had ED YM IR and YC up at the higher level, and then I dropped for NA and UI."
Own opinion	Where participants expressed a thought on their own opinion	"I would probably go for that but…if you put me in that situation where I had it and then asked me to fill out that booklet I probably would be very different…I don't know either way whether I would want to those extra years or whether I'd be like oh no this really is much worse than I could imagine."
Desirability of health states	Where a participants discussed a specific health state	"I've got NA on zero, because I would I would be dependent on someone else if I needed help with washing and dressing and all my usual activities and walking."
The MCDA exercise	Where a participants discussed the MCDA exercise	"Yeah I agree with what's been said really, I think that, the the idea of a burden is not good one because if you uhm if you pay into a collective system, then there's an expectation that you get something out of it at some stage, and that's what we do we pay in. So I don't see the idea of burden as being relevant."
The valuation process	Where participants discussed the TTO or ranking exercise	"It says sort of 20 years here so that's the time scale..I would definitely agree with that I think there's a huge difference between short term and long term and yes as you say it can have a far greater effect in the short term"
Experience shared	Where participants shared their previous experience	"It does have a social impact though, you know it did on me. I mean friends got fed up waiting for me…"
Off topic	Comments that did not have a relation to the valuation of health states	"I have donated my body to the university for whatever they wish to do with it"

Table 7.1 Themes found during group meetings, along with definitions and example quote

7.2.2 Themes that could explain the effect of reflection and deliberation

The themes were reviewed to assess whether they may be relevant in explaining preference changes. The relevant themes were included for further investigation. The reasons for including or excluding themes are listed in Table 7.2. The theme of 'Agreement or disagreement' was included because it provides insight into the deliberative exchanges that took place between participants. The theme 'own opinion' was included because it provides information on instances of participants reflecting on their own beliefs and preferences. The themes of 'grouping of health states' and 'desirability of health states' were included because they reflect participants' views on the health states. The views that participants heard from other participants may explain why changes did or did not occur. Finally, the theme 'the valuation process' was included because it reflected participants' views on the process of health state valuation and it may be that preference changes are due to changes to how people perceived the valuation task.

Several themes were excluded. The theme 'the MCDA exercise' was excluded because the comments covered specific aspects of the MCDA exercise but did not indicate why TTO changes would take place. The theme 'experience shared' was already covered in 'desirability of health states'. Finally, off topic comments were excluded, as they were not relevant to how participants valued health states.

Themes	Included or excluded	Reason
Agreement or disagreement	Included	Indicates the type of exchange between participants
Grouping of health states	Included	Indicates view of states
Own opinion	Included	Indicates reflection
Desirability of health states	Included	Indicates view of states
The valuation process	Included	Indicates point of view about health state valuation
The MCDA exercise	Excluded	Does not reflect TTO changes
Experience shared	Excluded	Covered by point of view about health states
Off topic	Excluded	Does not reflect TTO changes

Table 7.2 Themes and reason for including or excluding themes

7.2.3 Summarising themes

The five included themes are now summarised, using either a qualitative narrative summary or by quantifying the qualitative data, if quantification allowed for better presentation and summarising of the theme. Example quotes are provided in for each theme.

Theme 1: instances of agreement or disagreement amongst participants

This theme covered all mentions of participants agreeing or disagreeing with each other. Though there were instances where participants agreed with each other this was not necessarily illuminating in explaining the effect of reflection and deliberation. But the aspect of this theme that may explain changes was the disagreement amongst participant. The disagreements amongst participants could be categorised as 'reaching agreement', 'agree to disagree', or 'ignored'.

Sometimes participants reached agreement so that at least one participant changed their mind. In the following quote the participant acknowledges and agrees with the arguments from the other participants:

Yeah I didn't really think about the anxious and depressed really, I just thought like what would other people think I was like. And if someone was depressed they might not know that, so that's probably why I was thinking, dignity didn't really have a big role in anxious and depressed, but now they said it I kind of agree with what they said. (Group 6, P3)

Sometimes participants 'agree to disagree' meaning that participants acknowledged but maintained their disagreement. The following dialogue illustrates how participants maintain their different points of view on the effect of health problems on personal relationships:

P1: Can I just say I didn't agree with that, because I think people can have extremely good personal relationships but have all these physical problems....well uhhmm, well I'm thinking about say a child with cerebral palsy something like that, he might have severe problems walking about, they might have severe problems washing and dressing themselves, they might not be able to do the usual activities that other children can do. They might have pain and discomfort at times but they might have wonderful....relationships with their parents.

P3: See I have a total different experience a couple years back when I was ill and....uhh you didn't see anybody for dust. Both friends and family. (Group 2)

Finally disagreements could also be 'ignored' if the disagreement was stated but the next comment was not related to the disagreement. In the following dialogue the third speech turn moves away from the disagreement over the effect of anxiety and depression on being a burden to the definition of burden and whether it includes wider society:

P1: I didn't really think about being anxious and depressed as being a burden on other people, cause you don't have to be a burden on other people, you don't HAVE to have people...if you lived on your own and you didn't, I don't know, say you didn't go to work, you wouldn't, you don't have to rely on someone else to do something for you if you're in that state. Whereas if you can't wash or dress then you do and that is then I guess, that's more of a burden.

P4: I think sometimes if you're anxious or depressed you can't, you can't make any decision at all, you can't think straight

P5: I think I was thinking as well about being a burden on society sort of thing (...) (Group 3)

Each case of disagreement was coded as one of the three categories. There were a total of 33 comments that were classified as disagreement. There were disagreements present in all groups (see Appendix 17). The majority of disagreements were classified as 'agree to disagree' or 'ignored'. Of all instances of disagreements during the group meetings 19 (58%) were classified as 'agree to disagree', 9 (27%) as 'ignored', and 5 as 'reaching agreement' (15%). There is thus little evidence that deliberation was a process of solving contradictory beliefs. The evidence rather suggests that participants maintained their own opinions or ignore disagreements with other participants.

In cases where participants did not resolve disagreements, the disagreements were sometimes explained away due to personal circumstances (such as personality, family support, or age), or past experience with ill health. The following dialogue illustrate how participants agreed to disagree because of the different personalities of participants:

P6: I ranked YM higher...than YC...and that's because...uhhh...firstly there's problems with walking about (...) I'd like to do lots of photography and if I'm not able to walk (...) then obviously this would take a hit and so it was...the fact that I would have moderate problems walking about and the fact that my usual activities are outdoors that made me decide.

P4: I did the opposite because I thought that you could, I would rather do my usual activities and get there like...like by public transport or some different way. I don't really walk that much, so I could rely on other things really and still be able to do my usual activities, which I prefer. Which I I yeah, walking is not really an activity which I'd like to do anyway.

P1: That's fair because if you were a big walker then probably walking would have an impact, but if you like sitting and watching DVDs, then not being able to walk around might not be that much of a problem. (Group 3)

The dialogue below illustrates how family circumstances were used to explain disagreements:

P3: I was just going to say one more thing as well, I was going to say, I I I've never been in the situation, so maybe if I was I change my mind, but I think I disagree with you because I think that I'd be more kind of embarrassed if I had to get other people who I know who would uhm to kind of come and help me. I'd rather get someone else and be presentable and dressed when I met people I know.

P1: You might think different when you get married.

P3: Haha yeah possibility.

P1: I wouldn't, I don't know, I've got a very good relationship with my husband and I've looked after him because he had problem with his back and I did look after him and he's alright now and I know he would do the same for me. So again it's having that relationship that good relationship. Support, family support. (Group 5)

Overall, the deliberation is not necessarily characterised as solving contradictory beliefs about health states or by learning from other participants. Participants did not solve their disagreements but accepted them as being due to personal circumstances or simply ignored them. Preferences over health states often seemed to be regarded as being based on personal judgement and circumstances.

Theme 2: comments about the valuation process

Participants emphasised the personal aspect of the valuation process. In total 39 comments were made about the personal nature of health state valuation and these comments were present in all groups (see Appendix 17). Participants indicated that they considered previous experiences and personal circumstances to value a health state. The following three quotes illustrate participants explaining their health state values based on their previous experience with ill health, personal circumstances, and family situation:

I actually ranked extreme pain uhm as being the worst of uhm conditions. I think partly because one of my sisters suffered from a particularly chronic condition and just watching her suffer was pretty bad. I haven't really experienced myself or through friends uhm mental health problems, but extreme pain I have I think with my sister. So yeah so extreme pain scored very badly for me. (Group 7, P3)

I think like, it depends on your current age as well, I took that into account. Cause it says you live for another 20 years I'm only [20s] now, so I don't have any kids or anything, so my circumstances are different. (Group 5, P3)

I think also it depends on, there again, the attitude of your family. Uhm, my daughter is confidently expecting that her parents will become aged decrepitates and need to be looked after. And uhm, she has plans apparently hahaha. But my attitude to being a burden sort of reflects her's and our son's as well. (Group 8, P4)

Adaptation was one aspect that was often discussed during the group sessions. The ability to adapt to health problems was considered to depend on personal experience and personality:

I'm thinking about other people's experiences as well, so you can see what how other people are going through (...) I've seen enough people who've experienced levels of pain that I would think (...) oh if I was in that much pain I'd just want to prefer to die but then you see other people coping with it and you realise actually (...) there's different sort of things that are important. For some people dignity is incredibly important (...) it's about whether other people's perceptions will affect you. So you have to try to work out which group am I in? (Group 4, P2)

Two participants summarised the idea of the personal nature of health state valuation:

Also your own experience, so I've said like I've said, I've had a son with depression then I have a certain...view on that. Because I've got an elderly mother then there are other factors that come into play. They'll be for all of us, different uhm, different focus wouldn't it? Because of our personal experience. We're not doing this...we're doing this subjectivity really aren't we? (Group 2, P3)

Well one of the difficulties about all this sort of things I think is that, we already to some extent we've said it there's the health state are intensely personal, aren't they? So you could be in a condition I consider to be wretched but you could be having a reasonable quality of life and vice versa. (Group 4, P1)

These quotes illustrate that participants' health state values were seen as personal, shaped by their own experience and circumstances, which are not necessarily shared amongst participants. There was little evidence that the deliberation was a process of an

objective fact-based knowledge exchange, where novel information was shared and factual beliefs 'improved'.

Theme 3: comments about own opinion

This theme covers self-reflective comments. There was little indication that participants changed their minds or indicated surprise about hearing new information from other participants during the deliberation period. In total, there were only 11 mentions of participants changing their mind. This did not occur in all group meetings (see Appendix 17). Many of those comments were about slight changes to MCDA scores due to errors in scoring, rather than changes in preferences or perception of health states. Two examples of participants indicating a change in mind are illustrated in the following quote:

> *P4: when I first did it I uhhhh I ranked UI higher than YC and now after like being talking...I think that should be the other way around.*
>
> *P2: I think if we all did this again, took this off and gave it to us again tomorrow we'd probably all put different numbers heheh, we'd be going over in our heads subconsciously hehe and if we came back and looked at these again in a day or two we'd we'd have all this discussion going around in our head and it would change our views probably. (Group 1)*

There were only a total of three references amongst all the groups about participants not having considered adaptation, for example:

> *P2: ...whereas I feel that there are aids around to help you with the with your walking difficulties.*
>
> *P1: Hmm I never thought of that, I just thought of being able to go out and get a bus in the morning and walk about but yeah, being independent, never thought about aids yeah. (Group 10)*

There were three occurrences of participants being surprised or not having considered something, for example:

P2: So I used, whether there was any anxiety or not as a sort of final criteria for me in how I rated.

P1: See again I never thought about that aspect at all, I was just thinking of the physical side, of somebody having to wash you and wipe your bottom like you said and all these kind of things, all those kind of pressures, but I didn't really think of the anxiety. (Group 10)

P1: I find it worrying when you say about people screaming in the extreme pain, because out of ignorance I tend to think pain is controlled.

P5: Unfortunately no.

P1: Oh dear! (Group 2)

There was, however, a tendency for participants to indicate uncertainty in their valuation. There were various reasons for this uncertainty: the health state was too abstract, lack of experience, uncertainty about values or judgements, and difficulty of completing a TTO. Overall there were 47 instances, at least one in 12 of the 13 groups, where a participant mentioned uncertainty. An example quote indicating the abstractness and difficulty with valuing the abstract state is below:

(...) but it's hard to imagine, unless you've had those problems or know somebody that's got them I guess it's hard to imagine what slight problems might be and what scale that falls under (Group 3, P1)

Some participants expressed an uncertainty about whether their choices would be the same if they had experienced the health state:

I would probably go for that but then I, I think if you put me in that situation where I had it, and then asked me to fill out that booklet I probably would be very different. And I don't know, I don't know either way whether I would want to those extra years or whether I'd say oh no this really is much worse than I could imagine in this situation. (Group 1, P2)

A more general sense of uncertainty about a participant's own judgments is expressed in this quote, where a judgement about the physical health state is followed by a declaration of uncertainty:

(...) but you are more mobile...and therefore your personal, other aspects of you and your personal relationships are better. To some extent. I think. I don't know. (Group 6, P5)

The following quote is an illustration of difficulty in making choices due to the difficulty of the TTO:

Partly what made the first questionnaire [TTO] so hard was trying to evaluate the impact these things [personal relationships, enjoyment, and other criteria] would have in terms of the number of years you had to suffer them and that that was actually very difficult to tease out. (Group 11, P3)

Overall, there was a lack of evidence that upon reflection participants stated that their beliefs or preferences were changing or that they encountered novel information. Yet, participants often expressed uncertainty about their beliefs and preferences.

Theme 4: comments about grouping of health states

There was evidence that during the valuation process the participants grouped similar health states together, dividing health states in a group of 'good' and 'bad' health states:

But what I found is the polarisation, which is some them I would put very near to the top and some that I would put very near the very bottom of the scale. But it's difficult to pick things that I would put near the middle. (Group 2, P1)

For me they fell into two groups, ED, YM, IR, and YC, were all the OK end. And GY, AU, NA, UI, were all the not OK end. And really apart from that, you know you can I can give one a bit of a higher score than the other one but for me it really polarised into those two groups. (Group 6, P5)

Within the higher valued group, the two health states 11331 (labelled YM) and 31131 (labelled YC) were often valued similarly:

I put 70 for two of them. I put 70 for YM and YC. Because they in general, you are still able to...operate. You know, you're still able to do normal things, although there may be some pain or some moderate problems, they didn't seem too restrictive to me. I suppose, maybe I've had some of these problems before short term, or in a short-term ill health, and I haven't felt too worried about them. (Group 8, P5)

The health states of 44553 (labelled NA) and 44535 (labelled GY) were also valued similarly:

I scored GY and NA low down in the dignity scores. You can be very disabled in every way and still maintain your self-esteem, your dignity, whatever word you want to use for it. But I think when the the the when you can't look after yourself physically with severe problems walking bout, severe problems washing or dressing and load everything else on top of that as well, I think that's really when your dignity gets hit. (Group 7, P1)

This theme occurred in all group meetings (see Appendix 17). This theme provides some evidence that participants focused on the similarities of the health states and disregard the differences, therefore there was a polarisation of health states in two groups, the 'good' health states of 11331 and 31131 and the 'bad' health states of 44553 and 44535.

Theme 5: comments about relative desirability of the six health states

This theme describes how favourably participants described the health states. The data were first quantified and the analysis was conducted at the aggregate level for all groups because of the similarity between groups (see inter-quartile range shown in Table 7.3).

All comments about health states were categorised as positive, neutral, or negative mentions. An example of a positive mention is the following quote:

My second one was YM (...) because I'm quite happy, I mean I'm quite happy to put up with a bit of pain and uhm I'm sure I can overcome my moderate problems, with the usual activities, so it's not gonna make all that much difference to the life style. (Group 11, P2)

An example of a negative quote is the following:

IR (...) I just felt that everything else was you know, you have...problems within each of these domains and I felt that would probably build up and impact on how you perceive your dignity (...) I just think that it builds up and then you know feel frustrated that ok not only may have slight pain but I can't do the things that I want to do. I'm a little bit, there's a bit taken away from each aspect of your life. (Group 1, P5)

The quantified summary of this theme is reported in Table 7.3. For each health state the number of negative comments was subtracted from the number of positive items

and this number was ranked from best to worst. The ranking of the health states in order of most positively discussed is: 11331, 31131, 32322, 11334, 44553, and 44535.

Health state	**Average Count of Positive mention per group (IQR)**	**Average Count of negative mention per group (IQR)**	**Positive minus negative**	**Ranking**
11331	5.3 (5, 6)	0.8 (0,1)	4.46	1
31131	5.5 (5, 6)	1.2 (0,2)	4.38	2
32322	1.5 (1, 2)	1.2 (0,2)	0.23	3
11334	0.4 (0, 1)	4.6 (4,5)	-4.23	4
44553	1.0 (0, 2)	5.5 (4,6)	-4.54	5
44535	0.0 (0, 0)	7.6 (6,9)	-7.62	6

Table 7.3 Summary of theme comments about health states, states are ranked by average favourability

Themes	**Summary**	**Hypothesis**	**Quantitative findings**	**Comparison**
Agreement or disagreement	Health state valuation is largely personal. Disagreements are generally not solved but are explained away	No systematic aggregate changes; but possible individual level changes	No statistically significant changes at the aggregate level. Relatively large number of changes (61% to 82% of valuations for each health state) at the individual level.	True at both levels, concordance
The valuation process	Health state valuation is largely personal; adaptation, environment and context of valuation are all seen as personal			
Own opinion	Few indications of change of mind or being surprised, but uncertainty about opinion			
Grouping of health states	Health states are polarized between 31131 and 11331 compared with 44535 and 44553	At individual level health states 31131 closer to 11331 and 44535 to 44553	About 50% of individuals who made changes had more similar values for 31131 and 11331 after than before; same for 44535 and 44553	Discordance
Health states	Health states ranked by favorability during discussion: 11331, 31131, 32322, 11334, 44553, 44535	Second TTO should conform to ranking. In particular the ranking of 31131 and 11331 should reverse. Same for 44535 and 44553.	Rankings conform except for that 11331 is not ranked above 31131	Mixed. Concordance for 44553 compared to 44535, but not for 11331 compared to 31131

Table 7.4 Summary of themes, hypothesis generated from summary, related quantitative findings, and comparison between qualitative and quantitative findings

7.2.4 Developing quantitative hypotheses

This section aims to develop quantitative hypotheses based on the findings described in the previous section. The qualitative evidence of the group data suggests that:

- Participants tend to ignore disagreements or to 'agree to disagree'
- Health state preferences are considered personal
- Participants are uncertain about their values but hear little surprising or novel information from other participants
- Health states 11331 is viewed similarly to 31131 and health states 44553 to 44535
- The rankings of the health states in order of most positively discussed is: 11331, 31131, 32322, 11334, 44553, and 44535

The first three points would suggest a lack of large systematic changes in people's valuations. However, point three suggests that there may be changes at the individual level because participants are uncertain about their valuations. Therefore, there may be individual changes but they may cancel each other out in larger sample.

Point four would suggest that the TTO values of 11331 and 31131, as well as 44553 and 44535, should be closer together at the individual level in the second than first TTO as a result of participants grouping these health states into 'good' and 'bad' blocks. By putting the states into the blocks, the participant ignore their differences and focus on their similarity and thus value the states within the same block the similarly. As a result of the individual level changes, the aggregate TTO values should be closer in the second TTO than in the first.

There is a difference between the ranking of the first TTO and the ranking based on the deliberation period. When compared with the ranking of the first TTO (see Table 6.4) to the ranking of the discussion, the position of health states of 44553 and

44535 are reverse with 44535 being rated higher in the TTO. The position of health states 11331 and 31131 are also reversed with 31131 ranking higher in the TTO. Therefore, point five would suggest that the states 44535 and 44553 should reverse in the second TTO, with 44535 being ranked below 44553. As well, 31131 should rank below 11331 in the second TTO. If the type of discussion influences the participants, the expectation would be that a larger number of people agree with the ordinal rank of the health states as they were discussed in the deliberative period.

7.2.5 Testing hypotheses using the quantitative data

The hypotheses can now be compared to the quantitative data. The hypotheses and findings are shown in Table 7.4. Hypothesis one indicates no systematic aggregate changes but the possibility of changes at the individual level. Quantitative findings indicate little systematic change at the aggregate level because no statistically significant changes were found at the aggregate level and the average absolute change over all six states was 0.03. There are, however, more changes at the individual level and many participants made changes to their original health state values. Thus both at the individual and aggregate level the evidence from the two components here show concordance.

Hypothesis two indicates that the values of health state 31131 and 11331, and 44553 and 44535 should be more similar in the second TTO than in the first. This was not found in the quantitative component. Only half of individuals had more similar values for both pairs and about half the participants had larger differences between the states. The quantitative findings indicate that the states did not become polarised into groups that made participants ignore differences of the health states between the health states. The qualitative and quantitative finding show discordance.

Hypothesis three indicates that the ranking of the second TTO should be different from the first; in particular the state 44553 should be ranked above 44535 and the health state 11331 above 31131. Health state 44553 was indeed ranked above 44535 at the aggregate level in the second TTO. At the individual level, there are more participants in the second TTO that agree with the ordinal ranking suggested by the discussion for 44535 and 44553 than in the first TTO (the number of participants who ranked 44535 above 44553 decreased from 30 to 25 and the number of participants who ranked 44553 above 44535 increased from 16 to 22). The predicted change in ranking between 11331 and 31131 did not occur and health state 11331 was not ranked above 31131 (although the values differ by only 0.01). At the individual level the number of participant who ranked 11331 above 31131 did not change, and two more participants ranked 31131 above 11331 than before.

The qualitative and quantitative findings for this hypothesis show concordance for 44535 and 44553 and discordance for 11331 and 31131 and therefore the overall results are mixed. The views of the desirability of the health states expressed during deliberation thus did have some ability to predict changes in the second TTO, but only in one of the two pairs of states.

7.3 Discussion

7.3.1 Summary

Two possible reasons mentioned in the literature for why deliberation may be useful are to reveal information not known to all participants and to reduce mistakes in reasoning (Fearon, 1998). Both these reasons did not seem applicable to the findings in this study. The qualitative analysis suggests that information revealed to the participants was not generally noted as being surprising or novel. Deliberation could not be characterised as reducing error because rather than focusing on objective knowledge

exchange it focused on exchange of personal subjective beliefs. Participants focused on personal values, circumstances, and experiences and not on whether theirs view were correct or incorrect. In cases of disagreements the participants did not generally accept other participants' opinions. There was similarity here to study one (see Chapter 5) where conversion factors, some of which are personal, affect how individuals value health. At the same time participants were uncertain about their values and beliefs. Together the findings in this chapter suggest that reflection and deliberation, as implemented in this study, is unlikely to cause large systematic changes but changes at the individual level can be expected because individuals remain uncertain about their values.

The degree to which deliberation contained positive or negative comments about health states predicted one change in the ranking of the first to the second TTO valuations. This can be thought to be inconsistent with the conclusion that systematic changes were deemed. One possibility is that that large systematic changes are unlikely but that marginal changes are possible. Indeed sometimes participants did come to agreement and some participants did mention being surprised by information heard in the group meetings, but this did not occur frequently during the deliberation period of the group meetings.

7.3.2 Comparison to the literature

Two of the studies cited in the literature review of Chapter 3 on reflection and deliberation conducted a qualitative analysis in order to explain their results. Stein et al. (2006) conducted semi-structured telephone interviews with participants after all group meetings had been conducted. The findings from the interviews indicated that participants valued the discussion for procedural reasons: greater reassurance, improved procedural performance, increased group cohesion, and satisfaction of curiosity. The

authors did not suggest that discussion or hearing experiences of others changed beliefs but rather provided "reassurance about the individual's initial response". Any changes to an individual's health state values occurred because deliberation encouraged the participants to correctly implement the valuation task and not because deliberation changed an individual's true preferences. The authors did not analyse the group deliberation itself, but commented that their "general impression was that participants discussed their personal attitude to the scenarios and presented little new information to the rest of the group."

Robinson and Bryan (2013) also conducted interviews after the group meetings. The findings from the interviews suggest that the participants were unlikely to be influenced by other participants' valuations. Two factors may have prompted changes to the participants' preferences. First, the facilitator explored the implications of their responses to the PTO exercise in terms of preferring different group of patients. Second, each health state contained a disease label and discussion around those specific diseases took place.

The findings of Stein et al. (2006) resemble the qualitative finding in this study where the group meeting did not consist of deliberating on an objective knowledge basis but rather on sharing beliefs and experiences from a personal point of view. The findings in Robinson and Bryan (2013) do not resemble this study. First, they may reflect the difference between PTO and TTO task. The PTO task is based on a societal perspective and therefore discussing implications for trading people may encourage participants to consider equity implications. The TTO is based on a personal perspective and there are less direct equity implications. Second, this study did not use disease labels and that may have reduced the amount of information shared by participants beyond the description of the health state. It also meant that disease labels did not influence participants' initial valuations of the health states.

7.3.3 Implications for practice

The implication of this study is that reflection and deliberation may not be needed to be included in health state valuation tasks for the purpose of constructing preferences. It was expected that reflection and deliberation would give participants an opportunity to help decide on the conversion factors and consequences of the health states. However, it appears that the conversion factors and consequences are considered personal and that there is limited room for using other people's experiences. It could thus be argued that reflection and deliberation, as designed in this study, may not be needed to develop complete preferences and conventional methods, with one-off interviews, are not invalidated by the findings in this chapter.

7.3.4 Limitations

One limitation in this study is that participants' comments were taken as reflections of their thoughts, but participants may not act in accordance to what they say and may not express all their thoughts. For example, participants could initially disagree with others but may change their minds changed later. Participants maybe not express their surprise even if they are surprised. In addition, several stages of analysis were conducted to integrate the qualitative and quantitative components. The number of stages means that the findings are further removed from the data and what participants have said, therefore greater interpretation of the qualitative data was required on behalf of the researcher.

Some themes were summarised by quantifying qualitative data and each comment was weighed equally. This may be reasonable since the number of comments is large. Alternatively the researcher would have to judge the strength of comments but that may be very subjective.

In this chapter it was assumed that the qualitative findings could explain the quantitative findings where the qualitative and quantitative components showed concordance. Alternatively, it is possible that the concordance is a coincidence and that some third factor was the real cause of the quantitative data.

The actual group deliberation was analysed in order to not depend on recall by the participants. This had a drawback because few data are available on the reflection (i.e. the MCDA booklet) aspect of the group meeting. Although participants referred back to their MCDA answers during deliberation, most data in this chapter represents the deliberation and not the reflection period. Interviews could have been conducted after the group meetings to investigate participants' views on the MCDA booklet. More broadly, interviews in general would have given a chance to obtain the participants own views on the entire group meeting. Future research could combine both interviews and analysis of group meetings.

Lastly, in this study all the data of all groups were combined. A more detailed analysis could be conducted using the group level data. This was not the approach taken in this chapter because it was found that there was significant uniformity and similarity in the content of deliberation amongst groups. Therefore it was decided that aggregating all groups could be deemed appropriate.

7.3.5 Future research

An area for future research is to consider other methods for guiding participants in constructing preferences. One suggestion in the literature that is not fully explored is to compare MCDA methods to composite methods (Brazier et al., 2007, p.139). It was previously mentioned that MAUT (or more generally MCDA) is one method for guiding preference construction. The group meeting in this study did use MCDA but only as a guide for reflection and deliberation and MCDA valuations were not provided

to participants. A study could investigate whether preferences measured using MCDA are different from those measured by the TTO. Participants could value health states using MCDA and their valuations could be fed back to them. It's an open question whether MCDA criteria should be the consequences of health states or whether like the HUI (Feeny et al., 2002) the criteria should be the health state domains.

Another area for future research is to provide external information to the participants. A lack of new information or accepted information was noted in the group meetings. To address this, participants could be provided with information from outside the group. This was the approach taken by McTaggart-Cowan et al. (2011), who provided video clips where patients discussed how they adapted to health conditions. Participants could, for example, be provided with information on the six consequences for different health states.

7.4 Conclusion

This chapter demonstrated the application of a concurrent mixed method study to explain the effect of a reflection and deliberation exercise on participants' health state values. Although participants engaged in deliberation they tended not to agree with the views of other participants and to see health state preferences as personal. Participants did not seem to encounter much new or surprising information from other participants. There was an indication that participants were uncertain about their own values. The extent of positive or negative discussion around a health state predicted two changes in the ordinal ranking of the health states, but only of those changes occurred. Overall the findings suggest that although participants will deliberate with others on their health state preferences, the deliberation is not likely to lead to large systematic changes. The findings would suggest that the role of reflection and deliberation, as implemented in this study, in developing complete health state preferences is limited.

The findings of this study are both positive and negative for the validity of conventional choice-based methods. It is positive because it reveals that deliberation and reflection are not necessary to elicit preferences over health states. It is negative because participants maintained that they were uncertain about their opinions and valuations. The deliberation revealed that participants used their beliefs and past experience to value health states but that there was large amount of disagreements amongst participants, of which most were not resolved. That would suggest that it is important to assess whether participants' beliefs are on the whole valid. The following chapter assesses whether participants' preferences over health states are informed by assessing whether individuals are informed about the health states they are valuing.

Chapter 8

Are preferences over health states informed?

The objective of the third and last study of this thesis is to develop and implement methods to determine whether preferences over health states held by members of the general public are informed. Several authors have suggested that preferences used in welfare economics should be informed (Harsanyi, 1977; Hausman and McPherson, 2009). Similarly, in the health economics literature several authors have suggested that cost-effectiveness analysis should use informed preferences over health states (Gold et al., 1996, pp.99-100; Hausman, 2006; Brazier et al., 2007 114-116). It is important to assess whether preferences over health states are informed because members of the general public are asked to value states about which they generally have no experience (Brazier et al., 2005; Hausman, 2006)

It is not entirely clear from the literature what being informed means or what one should be informed about and how to test what preferences are informed. In Chapter 5 it was argued that individuals value health states by assessing the consequences of ill health for their life. It can then be argued that one aspect of having informed preferences over health states should be being informed about the consequences of ill health. This study attempts to operationalize a test of being informed by assessing whether members of the public are informed about the consequences of health states on their life.

No literature was found on whether preferences are informed (see the literature review of Chapter 3). There is a literature comparing valuations based on SWB to valuations based on the TTO (Dolan and Metcalfe, 2012; Mukuria and Brazier, 2013). Previous empirical work has shown that there is some evidence that members of the

general public overestimate the effect of some health problems on subjective well-being (SWB) (Dolan and Kahneman, 2007) and have difficulty assessing the impact of adaptation (Brazier et al., 2005). These studies have generally concluded that preference-based valuations emphasise mental health domains less than SWB valuation. Yet beyond comparison of preferences to SWB no literature was found on whether members of the general public are informed about health states they are valuing.

The methods section first provides an overview of the methods used to test whether preferences are informed. Then the datasets used and the method of analysis are described. This is followed by the presentation of the results, which includes an analysis of the relationship between health and several non-health consequences and a comparison of expectations of members of the general public to the actual experience of patients. Finally, the implications of the findings of this study are discussed, with an emphasis on the implications for choice-based methods of valuing health.

8.1 Methods

8.1.1 Design of study: how to test whether preferences are informed

In this study, preferences are said to be informed if the expectations of members of the general public about the consequences of ill health conform to the actual experience of patients[19]. For example, on average, the belief of the public about the effect of mobility problems on the consequence of enjoyment should conform to the actual level of enjoyment experienced by patients with mobility problems. Several stages were completed to test whether expectations conform to experience. First, the expectations of members of the general public concerning the effect of EQ-5D health states for the

19. The word 'patient' in this chapter refers to people who are in some sub-optimal health state, rather than necessarily a recipient of health care services. The patient is in a health state that is being valued, whereas members of the general public are only imagining or remembering health states.

six consequences identified in study one (activities, enjoyment, independence, relationships, dignity, and avoiding being a burden) were estimated. Then, the actual experience of patients for those consequences was estimated using secondary data. The expectations and the actual experience were then compared. If members of the general public are informed there should be no or little difference between the expectations of the public and the experience of patients.

Measuring expectations

A measurement tool is needed to measure the expectations of members of the general public. In study two (see Chapter 6) participants were asked to reflect on six consequences of ill health using a MCDA booklet. As part of the MCDA booklet they scored health states on six consequences on a scale from 0 to 100. These scores are used in this study as the expectations of members of the general public on the consequences of ill health.

Measuring experience

The experience of patients is measured using secondary data that contains self-reported EQ-5D and questions about the same six consequences of ill health that were used to measure expectations. In the secondary data patients thus report their health state and their level of experience of at least one of the consequences. The experience of a patient in a health states can be predicted by analysing the relationship between the EQ-5D and the reported outcomes on the consequences in the secondary data.

The overall design of this study is shown in Figure 8.1. Each stage is described in more detail later in this section.

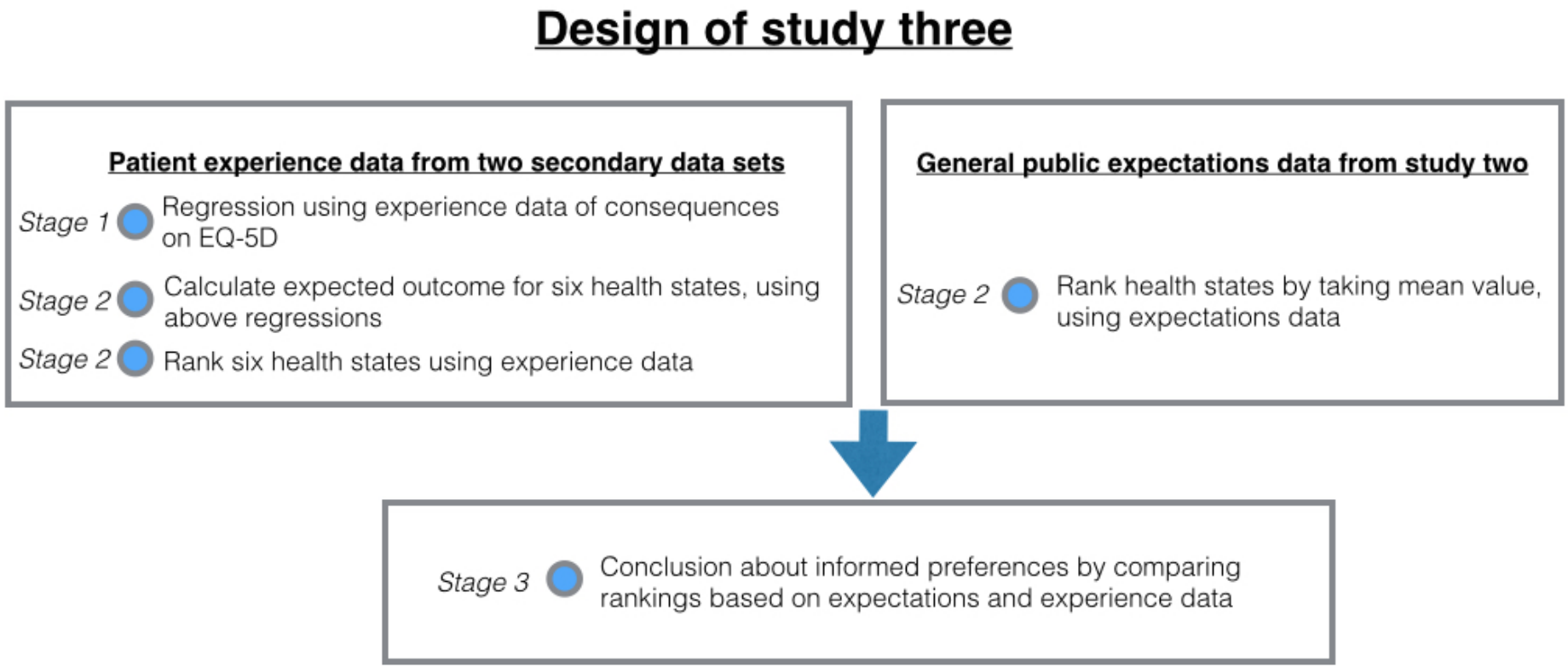


Figure 8.1 Design of study three

8.1.2 Datasets used

Three sources of data are used in this study. The first source is the data on expectations of members of the general public. The data was collected as part of the study described in Chapter 6 where members of the general public completed a MCDA booklet. The MCDA booklet asked the participants to score six health states, plus being dead and full health, on a scale from 0 to 100 on six criteria. These six criteria were based on the six consequences of ill health that participants of the study one (reported in Chapter 5) described as being important in health state valuation. The MCDA booklet is reproduced in full in Appendix 13. The six health states were those used in study two (reported in Chapter 6) and are reported in Table 6.1 (page 129). The dataset contains 62 participants with no missing data.

The second and third sources are the patient experience data. The two sources were chosen because they contained data on patients' experience on at least one of the six consequences of ill health and contained patients in a range of EQ-5D health states. Data for the dignity consequence were collected in a study testing generic preference-based measures for patients with pressure ulcers (Palfreyman and Mulhern, 2015). In that study dignity was measured using the Sheffield Dignity Questionnaire (SDQ) (Dixon, 2011). The dignity data contains 291 participants, with 265 complete cases.

Data for the other five consequences were collected in the Multi Instrument Comparison (MIC) study (Richardson et al., 2012). The MIC was a cross-sectional international study conducted to compare seven quality of life instruments and three instruments measuring SWB. The MIC study was conducted online in eight countries: Australia, UK, USA, Canada, Norway, and Germany. Participants in MIC were individuals with a self-reported disease diagnosis[20] (79% of sample) and a general public

20. Arthritis, asthma, cancer, depression, diabetes, hearing loss, and coronary heart disease

sample (21%). Online survey companies were used to collect data. The online survey companies approached their online panel until predetermined quotas of patients and a representative sample of members of the general public were recruited. The participants were first asked three subjective well-being questions (Richardson et al., 2015)[21]. They were then asked if they had any diagnosis of a list of diseases and if so, to select the most serious one. People allocated to a disease group completed a disease specific questionnaire before the main questionnaire (Richardson et al., 2015). The main questionnaire consisted of completing eight questionnaires: EQ-5D-5L, AQoL-8D, HUI-3, 15D, QWB-SA, SF-36, ICECAP-A, and a background questionnaire (Richardson et al., 2012; Richardson et al., 2015). Responses were deleted when an individual's completion time was less than 20 minutes. Several other rounds of data removal were made, based on inconsistent responses between the questionnaires (Richardson et al., 2012; Richardson et al., 2015). The MIC dataset contained data on five consequences and for a wide range of EQ-5D profiles, which made it ideal as a dataset to assess patient experience. The MIC dataset included 8022 participants, with 6776 complete cases.

Comparison of questions used to measure participant expectations and patient experience

Each question in the participants' expectations dataset is paired with one from the patient experience dataset but there are potentially important differences between their phrasing. A fuller discussion of the differences takes place in the limitation section of this chapter. The exact phrasing of the questions used to measure expectations, as well as labels for the 0 and 100 anchors, are shown in Table 8.1.

21. The Personal Wellbeing Index, the Integrated Household Survey, and the Satisfaction with Life Survey

Consequence	**Phrasing**	**Best anchor (100)**	**Worst anchor (0)**
Activities	Would you feel able to do the things and activities that you want to do?	Completely	Not at all
Relationships	Would you feel satisfied with your personal relationships	Completely	Not at all
Independence	Would you feel independent and in control of your life	Completely	Not at all
Burden	Would you be able to avoid being a burden on others	Completely	Not at all
Dignity	Would you feel you can maintain your dignity	Completely	Not at all
Enjoyment	Would you feel you are able to enjoy life?	Completely enjoy life	Completely unhappy

Table 8.1 Phrasing of questions and anchors for each consequence for expectations of general public

Consequence	Phrasing	Response levels (best to worst)	Dataset	Original source
Activities	During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your PHYSICAL health? Cut down the amount of time you spent on work or other activities.	1: None of the time 2: A little of the time 3: Some of the time 4: Most of the time 5: All of the time	MIC[1]	SF-36[2]
	(Same as above)…EMOTIONAL problems (such as feeling depressed or anxious)?	1: None of the time 2: A little of the time 3: Some of the time 4: Most of the time 5: All of the time		
Relationships	Love, friendship, and support	I can have a lot of love, friendship, and support I can have quite a lot of love, friendship, and support I can only have a little love, friendship, and support I cannot have any love, friendship, and support	MIC	ICECAP-A[3]
Independence	Being independent	I am able to be completely independent I am able to be independent in many things I am only able to be independent in a few things I am unable to be at all independent	MIC	ICECAP-A
Burden	How much of a burden do you feel you are to other people?	Not at all A little A moderate amount A lot Totally	MIC	AQoL-8D[4]
Dignity	Dignity	I feel that I live with dignity I feel that I live with some dignity I feel that I live with very little dignity	Dignity[5]	Sheffield Dignity Questionnaire[6]
Enjoyment	Overall, how happy did you feel yesterday?	11: Completely happy to 0: Not at all happy	MIC	Integrated Household Survey[7]

1: (Richardson et al., 2012), 2: (Ware and Sherbourne, 1992), 3: (Al-Janabi et al., 2012), 4: (Richardson et al., 2014), 5: (Dixon, 2011), 6: (Palfreyman and Mulhern, 2015), 7: (Office for National Statistics, 2012)

Table 8.2 Phrasing, response options, and source of questions used for patient experience data

The questions that were used for the patient experience dataset, along with their response categories, are shown in Table 8.2. The enjoyment question was originally developed for the Integrated Household Survey, which is administered by the British Office of National Statistics, and which measures happiness on an 11-point scale ranging from "Not at all happy" to "Completely happy" (Office for National Statistics, 2012; Hicks et al., 2013). One difference to note here is that the expectation question refers to enjoyment of life, whereas the patient experience question refers to happiness yesterday.

The questions for independence and relationships were originally developed for the ICECAP-A (Al-Janabi et al., 2012). Both are measured on a four-point scale. The independence consequence in the participant expectations includes the word of control, whereas the ICECAP-A phrasing only refers to independence. The relationship phrasing for the participant expectations mentions only personal relationship but the patient experience phrasing also mentions 'love, friendship, and support'.

The burden question was developed as part of the Australian Quality of Life measure and measures burden on a 5-point scale, ranging from "not at all" to "totally" (Richardson et al., 2014). The phrasing of this question in both datasets is very similar.

The dignity question was originally developed for the Sheffield Dignity Questionnaire and has three response levels (Dixon, 2011). The phrasing of this question across both expectations and experience dataset is very similar.

The activities question in the patient experience dataset was adjusted to ensure similarity with the expectations dataset, which asked participants to judge how ill health would limit their activities. For the experience dataset an activity limitation question from the SF-36 was used (Ware and Sherbourne, 1992). Unlike the expectations question, the SF-36 question separated limitations caused by mental and physical health problems. To obtain activity limitations due to both physical and mental health

problems, the responses to the SF-36 question were adjusted by selecting the worse response level reported between mental and physical health. For example, someone responding that they were "all the time limited" by their physical health but "none of the time limited" by their mental health was considered to be "all the time limited". The assumption is thus made that the total limitation caused by the physical and mental problems are equivalent to the worst response in either domain. The recoded question contained the same scale from 'none of the time' to 'all of the time'. The phrasing of this consequence in the participant expectations dataset refers to 'doing the things and activities you want to do' but the phrasing in the patient experience dataset refers to 'work or other activities'.

Difference in response options

The participant expectations data was collected using direct estimation on a scale from 0 to 100. The patient experience data uses rating scales, with between three to 11 response options. Therefore, the responses to the two datasets are on different scales. The response scales with three to 11 categories cannot be assumed to represent a continuous interval measure. As a result, only the ordinal placement of the health states on the scales can be used.

8.1.3 Analysis plan

To compare the participants' expectations to the patients' experiences several stages of analysis are conducted. Analysis consists of three stages after the summary statistics and the sample characteristics are reported. First, regressions are estimated to establish the relationships between the EQ-5D health states and the six consequences for the experience of patients. Second, the results of the regressions are used to predict the experience of patients for the six health states. These predictions are used to rank the six health states for each of the six consequences. The six health states are also ranked

using the participant expectations data. Finally, the experience and the expectations rankings are compared to see if they conform to each other. Each stage is presented in detail in the next sections.

All data analysis is done using R (R Core Team, 2015) and Excel. The R functions that are used are the VRRGLM function in the VGAM package (Yee and Wild, 1996) for the stereotype model, POLR in the MASS package (Venables and Ripley, 2002) for the ordered logit model, betareg in the betareg package for the beta regression (Cribari-Neto and Zeileis, 2010), and the LM function for the OLS model (R Core Team, 2015). For all regression analysis only complete cases with no missing data are used. This ensures that samples don't change with different regressions and allows for comparability across models.

Ethics approval was obtained for the use of secondary data (see Appendix 18).

8.1.4 Stage 1: regression analysis of patient experience data (for MIC and Dignity study)

First regression analysis is conducted to investigate the relationship between the consequences and ill health. This is done to estimate what the experience of a patient in a certain health states will be. The appropriate type of regression analysis depends on whether the consequence is measured as an ordinal or continuous variable. When ordinal, the response levels of the variable are ordered but the difference between the levels are unknown (Bryman, 2012, p.335). When continuous, the difference between the response levels are known (Stevens, 1946; Bryman, 2012, p.335). Five consequences in the patient experience data are treated as ordinal: activities, independence, relationships, dignity and burden. Enjoyment is treated as a continuous variable.

No background variables were used in the regressions because it is not possible to know what background characteristics participants were imagining when scoring the

health states. Additional analysis was undertaken which did include background characteristics in the regression but it did not change the results of the study (results not shown).

The regression models used for ordinal and continuous dependent variables are explained below.

Regression models used for ordinal outcomes

In the case of the ordinal variables a popular model in the literature is a cumulative logit model, also named the ordinal logistical regression (Ananth and Kleinbaum, 1997; Walters, 2009, p.258):

$$\log \frac{\Pr(Y \le y_i \mid x_1, x_2, ..., x_p)}{\Pr(Y > y_i \mid x_1, x_2, ..., x_p)} = \alpha_i + \beta_1 x_1 + ... + \beta_p x_p, i = 1, ..., k-1$$

where Y is the outcome variable, which has k categories represented by y_i, i = 1, 2,..., k and where x1, x2,..., xp are the p independent variables. The numerator is the probability of being in a category i or lower given the value of the covariates. The α_i parameters represent intercepts and are ordered as $\alpha_1 < \alpha_{k-1}$. This model is also known as the proportional odds model, because it makes the proportional odds assumption: the $\beta_1, ..., \beta_p$ regression parameters are the same for each response category i (Walters, 2009, p.259). The proportional odds assumption means the relationship between an independent variable and the dependent variable does not depend on the level of the dependent variable, therefore only one beta coefficients is estimated for each independent variable (Walters, 2009, p.259).

The proportional odds assumption can be tested using a graphical or a statistical test. The statistical test is conducted by comparing the cumulative logit model to a multinomial logistic model (MNL) using the chi squared score test (Bender and Grouven, 1997; Fox et al., 2011, p.271). Graphically, the test is conducted by estimating

a set of binary equations separately (rather than all at once as happens in the ordinal logistical regression) and plotting the estimated coefficients (Harrell, 2001, p.335). If the coefficients are similar, the proportional odds assumption is appropriate (Harrell, 2001, p.335).

In cases where the proportional odds assumption fails other models can be used. One alternative used in the literature is the multinomial logistic regression (MNL), which does not make the proportional odds assumption (Long and Freese, 2006, p.224). Neither, however, does it assume that the dependent variable is ordered. An alternative to the MNL model is to use a model that does assume the dependent variable is ordinal but does not make the proportional odds assumption. One such model is the stereotype logistic (STR) model (Walters, 2009, p.260). The stereotype model has recently become available in statistical packages and provides a compromise between the parsimony of the cumulative logit model and the relaxed assumptions of the multinomial model. The stereotype model is written as (Walters, 2009, p.261):

$$\log\frac{\Pr(Y=y_i\,|\,x_1,x_2,...,x_p)}{\Pr(Y=y_1\,|\,x_1,x_2,...,x_p)}=\alpha_i+\theta_i(\beta_1x_1+...+\beta_px_p),i=2,...,k$$

where Y represents the dependent variable with k ordered categories y_i, i = 1, 2,..., k. x_1, x_2,..., x_p are the p independent variables. The log odds for each y_i are formed by comparing that category to a base category y_1. α_i are the intercept terms. Similar to the proportional odds model, one set of β parameters are estimated per independent variable. However, the stereotype model also contains a set of θ parameters. θ_i are a set of parameters for each category of the dependent variable. θ_κ and θ_1 are constrained at 1 and 0 in order to identify the model, and the θ parameters are constrained to be ordered and thus the model will be ordered:

$1 = \theta_k \geq, \ldots, \theta_2 \geq \theta_1 = 0$

To obtain the coefficients for a level (Y=y_i), the θ parameter of that level and beta parameters ($\beta_p x_p$) are multiplied. The product $\theta_i \times \beta_p x_p$ is the log odds ratio for the covariate x_p of Y=y_i compared to Y=y1 for a unit increase in x_p. Thus the model assumes that the same θ parameter can be multiplied for each independent variable.

The stereotypy model is more parsimonious than the MNL model because it estimates one set of θ parameters instead of estimating a separate β for each level of the dependent variable. This simplification may result in a poorer fit, but this can be tested. One way to assess the appropriateness of the STR model is by comparing its predictions to the MNL model. If the predictions of the MNL and the STR are similar, the parsimonious nature of the STR would make it the preferred model (Long and Freese, 2006, p.282). In this study, if the proportional odds assumption is not valid and the predictions of the STR are similar to the MNL the STR model is used.

Interpretation of model results

The results of the STR model are most easily interpreted in odds ratio form (Long and Freese, 2006, p.283). Odds ratios are interpreted as the change in the odds of being in a certain response level compared to a base level, due to a unit change in the independent variable. Thus, for a continuous independent variable, the meaning of an odds ratio is the change in odds of being in level Y=y_i compared to the base level, due to a unit change in the continuous independent variable. For a categorical independent variable, the meaning of an odds ratio is the change in odds of being in a given level (Y=y_i) compared to the base level, for that independent variable level compared to the independent variable base level. In this study the chosen base level is the best response level on a consequence (e.g. no activity limitation). The base level chosen for the five EQ-5D domains variables is no health problems.

Regression models used for continuous dependent variables

The enjoyment consequence contains 11 response levels but is treated as a continuous variable. Walters (2009, p.255) recommends converting variables with seven or more response levels to a 0 to 1 scale and treating it as a continuous measure. Assuming that enjoyment is cardinal of enjoyment appears to makes little difference to model estimates for a set of background and economic variables[22] (Ferrer-i-Carbonell and Frijters, 2004). Assuming cardinality does avoid the problem of modelling ordinal data with a large number of response levels because a large dataset is needed to ensure that all 11 outcomes have sufficient data points. One possibility is to combine response levels to reduce their number but combining levels is somewhat arbitrary.

The enjoyment variable is transformed to a 0 to 1 scale using the formula (Smithson and Verkuilen, 2006):

$$y' = \frac{y - a}{b - a}$$

where y is the untransformed variable, y' is the transformed variable, a is the lowest response level, and b is the highest.

A beta regression is used to analyse enjoyment because the data is bounded between 0 and 1 and an OLS is not appropriate for bounded data (Smithson and Verkuilen, 2006). The beta regression model is able to model data bounded between 0 and 1. The beta regression assumes the dependent variable follows a beta distribution (Smithson and Verkuilen, 2006):

$$f(y;\omega,\tau) = \frac{\Gamma(\omega+\tau)}{\Gamma(\omega)\Gamma(\tau)} y^{w-1}(1-y)^{\tau-1}$$

where $y \in (0,1)$, $\Gamma(.)$ denotes the gamma function, ω and τ (>0) are shape parameters. The linear predictions are converted to a unit interval using the log link:

22. Age, income, living in partnership, number of children in the household, and health

$$\log \frac{u_i}{1-u_i} = x_i\beta$$

where u_i denotes the expected value of Y, x_i denotes the vector of covariates and β denotes the vector of regression coefficients.

The beta regression cannot handle values of 0 and 1 themselves, because the formula is undefined. A second transformation is applied because this dataset does contain 0 and 1 values. This transformation marginally compresses the values so that the model can be estimated (Smithson and Verkuilen, 2006):

$$Y^* = \frac{Y(N-1)+0.5}{N}$$

where Y^* is the transformed dependent variable, Y is the 0 to 1 bounded transformed data, and N is the sample size.

Inconsistent estimates from regression models

Once models are estimated, the regression coefficients are described. It is possible that estimated regression coefficients are inconsistent. Inconsistent coefficients means that worse health levels are associated with better outcomes on the consequences. This is arguably illogical. Therefore, for the EQ-5D domains with inconsistent and statistically insignificant coefficients the levels of the independent variable are combined until consistent coefficients are obtained. For example, where unable to walk about is associated with better relationships than severe mobility problems, and the coefficient for being unable to walk about is statistically insignificant, the two levels are combined. The adjustments are made first on the first domain of the EQ-5D-5L that had inconsistent coefficients and then made in the order of the EQ-5D questionnaire.

It was found that the levels four ("severe problems") and five ("unable to") of the mobility, self-care, and usual activities domains had inconsistent coefficients for five of the consequences (those collected form the MIC dataset). This may be due to the

limited sample size (less than 48 of the observations are in the fifth level). Therefore, those levels are combined for the MIC dataset models.

Stage 2: predicting the experience of patients and the expectations of participants in different health states

After the regression models have been estimated they are used to predict the experience of patients for the six health states on each consequence. The method for estimating predictions depends on the type of regression model used.

Predictions for categorical dependent variables

For the cumulative logit regression model probabilities can be predicted for of being in any response level for each health state. For the cumulative logit model they are calculated as (Walters, 2009, p.258):

$$\Pr(Y \leq y_i \mid x_1, x_2, ..., x_p) = \frac{\exp(a_{i+} \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p)}{1 + \exp(\beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p)}, j = 1, ..., k-1$$

For the stereotype model these are calculate as follows:

$$\Pr(Y = y_i \mid x_1, x_2, ..., x_p) = \frac{\exp(\alpha_i + \theta_i(\beta_1 x_1 + ... + \beta_p x_p))}{\sum_{i=1}^{k} \exp(\alpha_i + \theta_i(\beta_1 x_1 + ... + \beta_p x_p))}, i = 1, 2, ..., k$$

To obtain these predictions the EQ-5D variables in the model are set at the levels for the six health states.

The models predict probabilities of being in each level of the consequences for each health state. To get an overall expected value per health state the probabilities and response levels must be combined. This is done by multiplying the predicted probability by the response level and summing the total. The response levels are numbered from 1 to n, where n is the worst level. The expected value for each health state is used as a measure of the experience of an average patient in the health state.

For the dignity consequences the method is slightly different because the study used the EQ-5D-3L rather than the EQ-5D-5L (Palfreyman and Mulhern, 2015). Although the domains are the same, the EQ-5D-3L only contains three levels, while the EQ-5D-5L contains five levels. Therefore, predictions have to be made on the EQ-5D-3L states and then converted to the six EQ-5D-5L state of interest. The probabilities for converting EQ-5D-3L states to EQ-5D-5L states are reported by (van Hout et al., 2012).

Predictions for continuous dependent variables

For predictions using the beta regression model the formula below is used (Smithson and Verkuilen, 2006):

$$u_i = \frac{\exp(x_i\beta)}{1+\exp(x_i\beta)}$$

The EQ-5D levels are set at the values for each health state. The beta regression model will produce a predicted value between 0 to 1 for each health state. This value is the measure of the patient experience for continuous dependent variables.

Participants' expectations

The values for the expectations of the participants are calculated by taking a mean of their MCDA scores. This will produce values between 0 and 100. These values are the expectations of the participants for each health state.

8.1.5 Stage 3: comparing expectations and experience

Values for expectations of participants and the experience of patients are now available and can be to be compared in order to check for agreement. Comparing the expectations and experience can only be made on an ordinal basis because the two use different scales. Therefore, health states are ranked from best to worse according to each set of values. The two rankings are then compared to assess whether participants'

expectations of the consequences of health states are consistent with the experience of patients.

8.2 Results

8.2.1 Dataset description

Characteristics of samples

Respondent's socio-demographic characteristics for the participants expectation dataset from study two (reported Chapter 6) are presented in Table 6.2 (page 145). The median age (1st and 3rd quartile) is 44 (25 and 62) and the median EQ-5D-5L scores was 1 (0.77 and 1). About 59 percent of respondents are female.

Respondent's socio-demographic characteristics for the MIC and dignity datasets are reported in Table 8.3. The median age was 56 (42 and 66) and the median EQ-5D-5L value was 0.77 (0.66 and 0.88). About 52 percent of respondents are female. For the dignity dataset the median age is 77 (68 and 83) and the median EQ-5D-3L scores was 0.26 (-0.06 and 0.59). About 58 percent of respondents are female.

	MIC dataset	Dignity dataset
Number of participants	8022	291
Median age (1st quartile, 3rd quartile)	56 (42, 66)	77 (68, 83)
Median EQ-5D value (1st quartile, 3rd quartile)[1]	0.75 (0.66, 0.88)	0.26 (-0.06, 0.59)
Female, n (%)	52%	58%
Not married, n (%)	36%	60%
Degree, n (%)	28%	-
Full time employed, n (%)	33%	-
Part-time employed, n (%)	15%	-
Homemaker, n (%)	7%	-
Unemployed, n (%)	8%	-
Retired, n (%)	27%	-
Disability Pensioner, n (%)	7%	-
Student, n (%)	4%	-

1: Using the crosswalk UK value set (van Hout et al., 2012)

Table 8.3 Descriptive statistics for sample for MIC and dignity datasets

EQ-5D variable responses

The distribution of the responses on the EQ-5D-5L from the MIC dataset is shown in Figure 8.2. 'No problems' was the most frequently reported response for four of the five EQ-5D domains, while for pain or discomfort it was 'slight'. The 'unable to'/'extreme' level was rarely reported, ranging from 0.4% in Mobility to 2% in Anxiety and Depression. For mobility, usual activities, and self-care less than 1% of respondents reported the worst response level.

The distribution of the responses on the EQ-5D-3L from the dignity dataset is shown in Figure 8.3. The response level of 'some problems'/'moderate' was the most frequent reported outcome for all five EQ-5D domains.

Dependent variables

The responses on the dependent variables are reported in Figure 8.4. Median enjoyment was 0.7 (0.5 and 0.9). For all other five outcomes, the best response level was the most frequently reported.

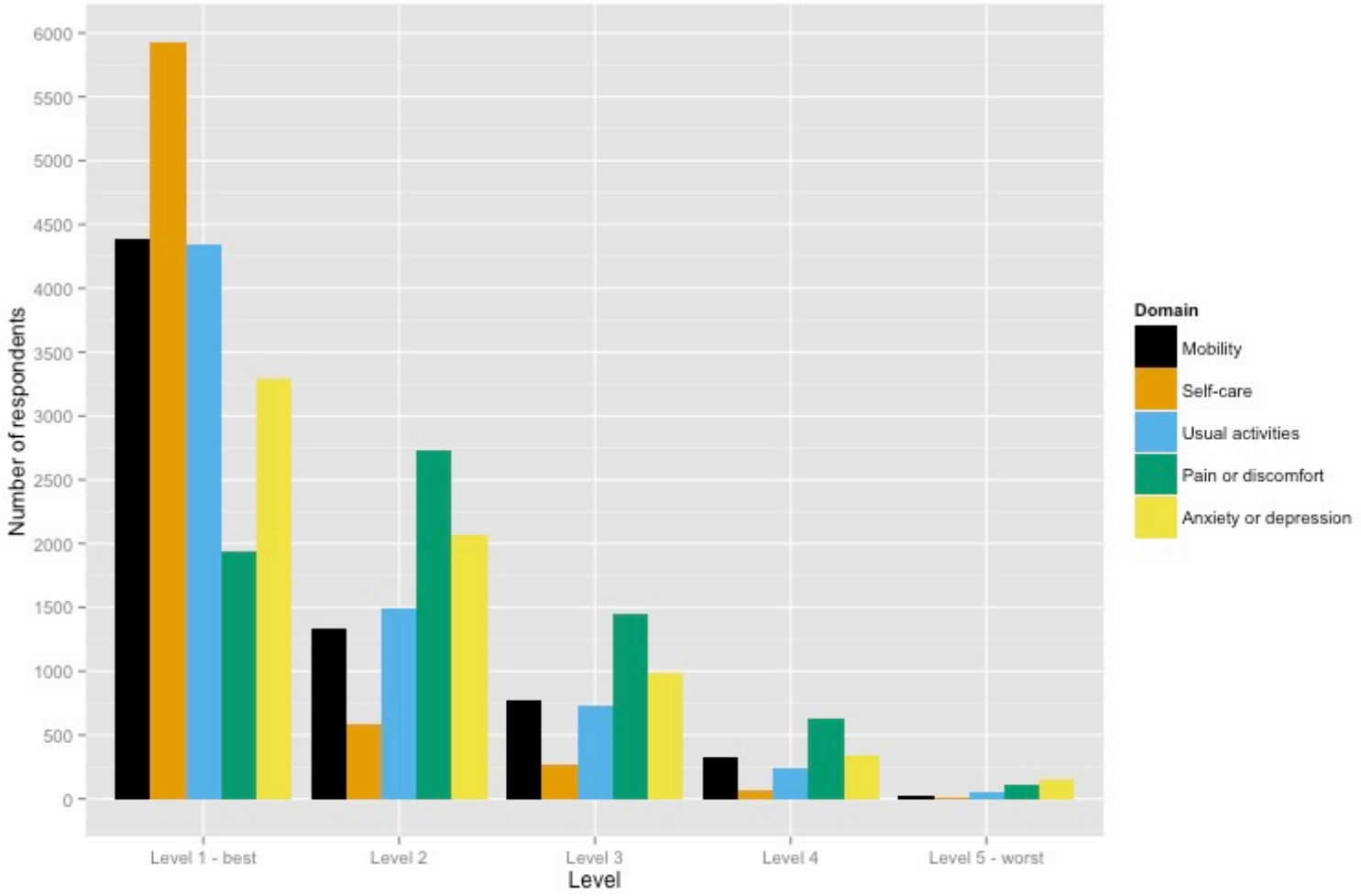


Figure 8.2 Number of respondents per level for each EQ-5D domain for MIC dataset

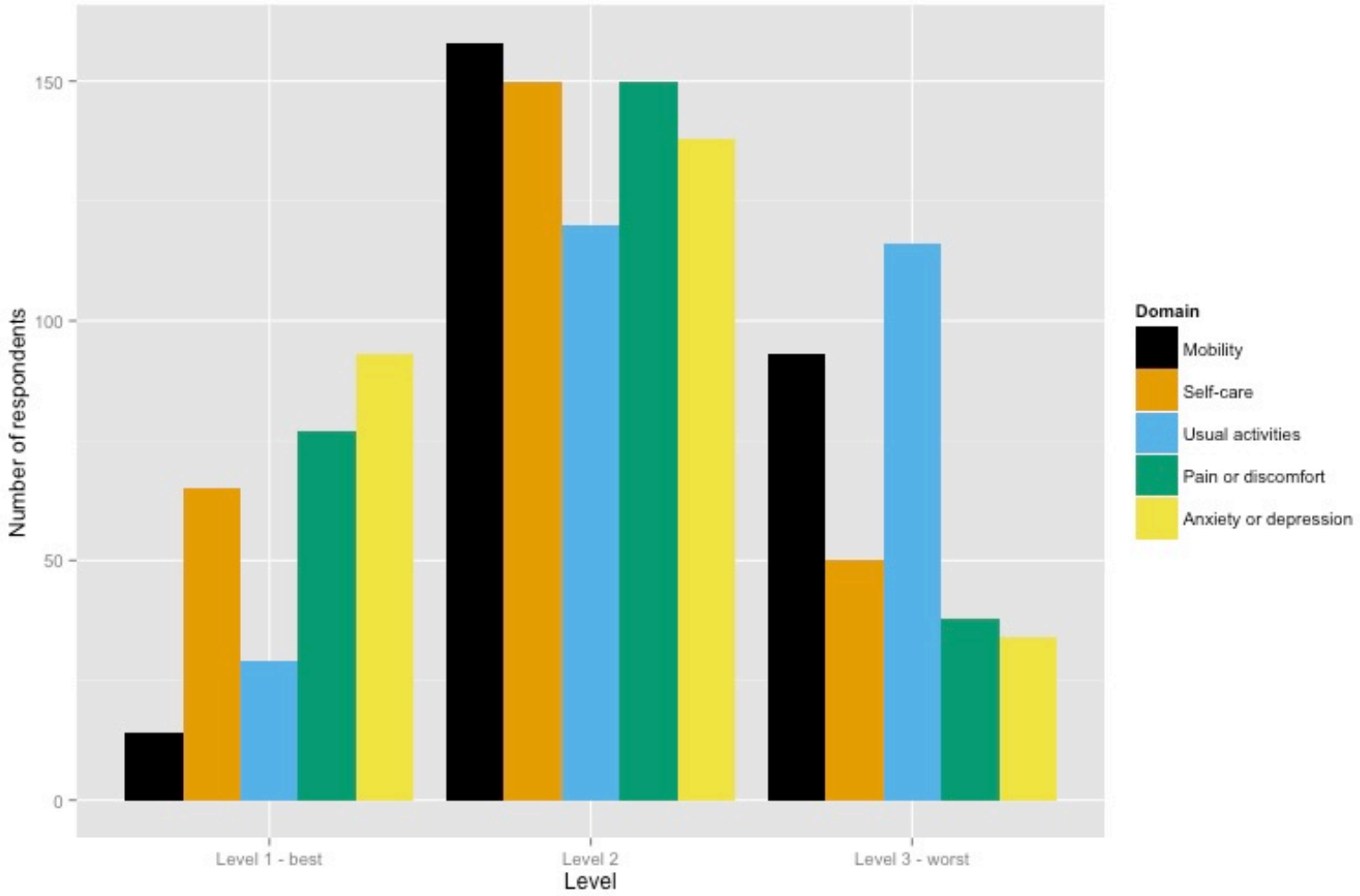


Figure 8.3 Number of respondents per level for each EQ-5D domain for dignity dataset

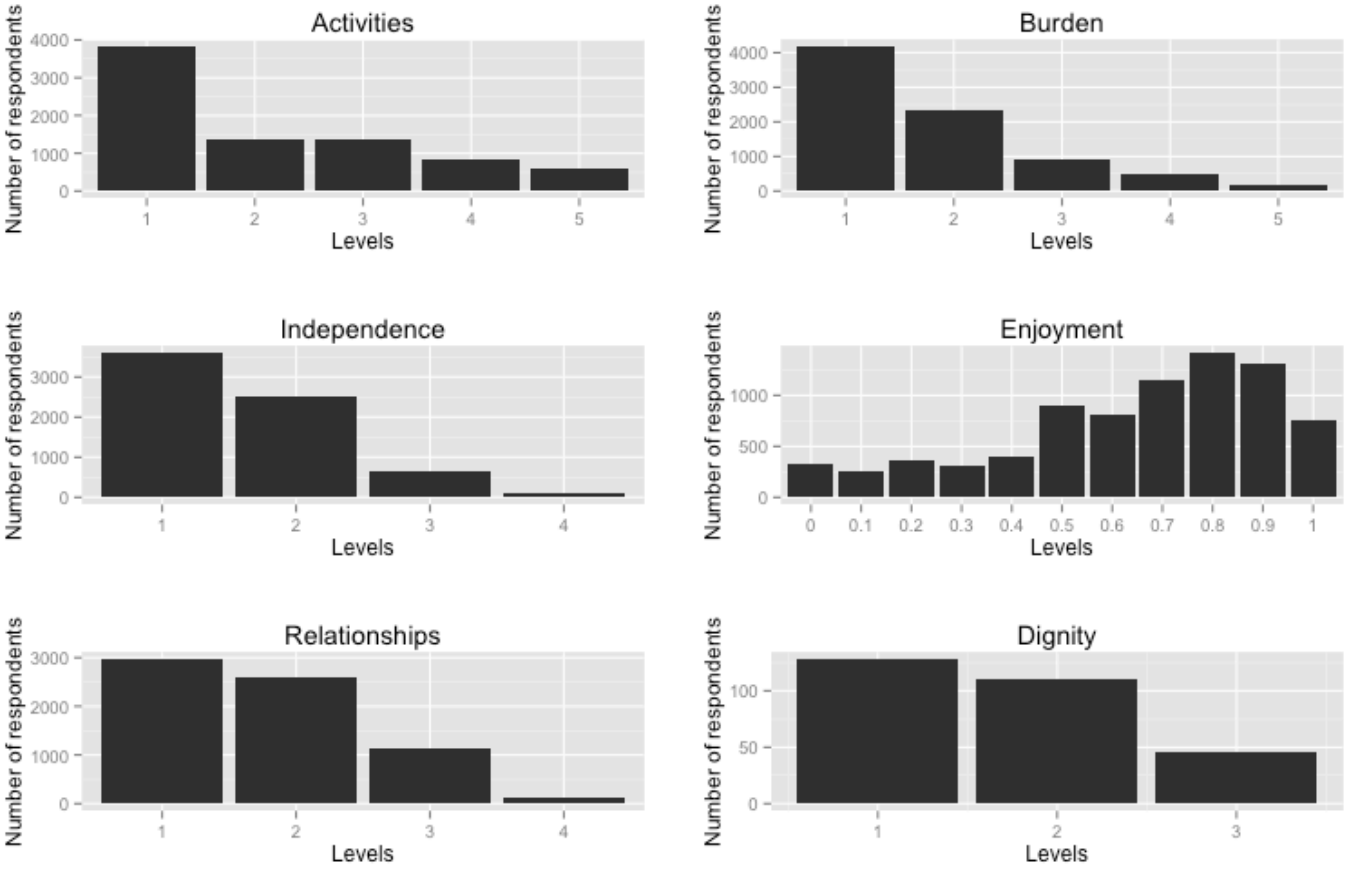


Figure 8.4 Distribution of responses on dependent variables. Level one is the best response level

8.2.2 Results of stage 1: regression analysis of patient experience data

The next six sections describe the regression results for each consequence. Each section provides a description of the chosen regression model, the variables included, adjustments made to obtain consistent estimates, and the estimated coefficients.

Consequence 1: activities

Using the ordered logit model, a chi square test comparison with the multinomial model shows that the proportional odds assumptions is violated (deviance of 347, with 51 degrees of freedom, p-value of <0.01). This is also shown graphically in Figure 8.5. If the proportional odds assumption was appropriate, the estimated log odds for each independent variable for each outcome level would be similar and thus the dots on each line would be close together (Harrell, 2001, p.336; Abreu et al., 2009). Therefore, the activities variable is also modelled using a stereotype logit model.

The full model is reported in Table 8.4. No inconsistent estimates were found and worse levels of health were associated with more activities limitations. The correlations with the predicted probabilities of the MNL model range from 0.87 to 0.99, indicating a high level of correlation. The final model is reported in Table 8.5 using the odds ratio form.

The highest odds ratio[23] for activities limitation is the usual activities domain with an average odds ratio of 25.5. The average odds ratios for the other domains are ranked as anxiety and depression (4.5), self-care (2.9), mobility (1.5) and pain (1.9). The size of the odds ratio indicates that the relative odds of having activities limitations increase the most when having problems in the usual activities domain.

23. Average odds ratio is the average of the odds ratios of a levels of a domain. It is used as a simplified method of comparing the odds ratios for an entire domain rather than per level of each domain.

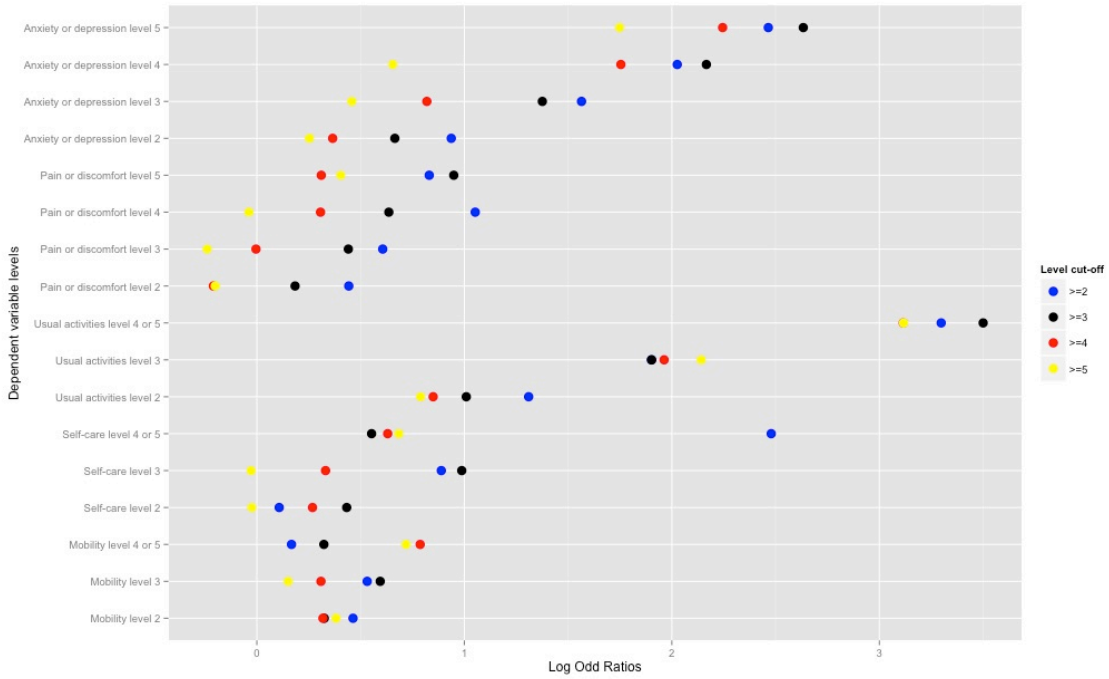


Figure 8.5 Proportional odds check for activities

Explanatory variables	**Coefficient (log odds)**	**Standard error**	**P-value**
Mobility level 2	0.63	0.11	0.00*
Mobility level 3	0.73	0.18	0.00*
Mobility level 4 or 5	0.89	0.35	0.01*
Self-care level 2	0.39	0.17	0.02*
Self-care level 3	1.37	0.41	0.00*
Self-care level 4 or 5	3.14	0.97	0.00*
Usual activities level 2	1.74	0.11	0.00*
Usual activities level 3	3.44	0.22	0.00*
Usual activities level 4 or 5	8.41	0.66	0.00*
Pain or discomfort level 2	0.52	0.10	0.00*
Pain or discomfort level 3	0.78	0.13	0.00*
Pain or discomfort level 4	1.39	0.21	0.00*
Pain or discomfort level 5	2.00	0.60	0.00*
Anxiety or depression level 2	1.26	0.09	0.00*
Anxiety or depression level 3	2.21	0.12	0.00*
Anxiety or depression level 4	3.39	0.23	0.00*
Anxiety or depression level 5	4.79	0.44	0.00*
Theta 5	1.00		
Theta 4	0.89		
Theta 3	0.71		
Theta 2	0.48		
Theta 1	0.00		
Constant 4	-4.99		
Constant 3	-3.90		
Constant 2	-2.55		
Constant 1	-1.84		
Observations	8021		
Log-Likelihood	9057		
AIC	18162		
Deviance	18114		
Correct predictive ability based on in sample prediction	51%		

* P-value < 0.05

Table 8.4 Full model for activities. Reference category is the best outcome level

Explanatory variables	Odds of level 2	Odds of level 3	Odds of level 4	Odds of level 5
Mobility level 2	1.35	1.56	1.74	1.87
Mobility level 3	1.42	1.68	1.91	2.07
Mobility level 4 or 5	1.53	1.88	2.19	2.42
Self-care level 2	1.21	1.32	1.42	1.48
Self-care level 3	1.94	2.65	3.37	3.94
Self-care level 4 or 5	4.54	9.30	16.08	23.07
Usual activities level 2	2.32	3.45	4.67	5.71
Usual activities level 3	5.26	11.56	21.08	31.32
Usual activities level 4 or 5	57.67	394.84	1714.24	4509.40
Pain or discomfort level 2	1.28	1.44	1.58	1.68
Pain or discomfort level 3	1.46	1.74	1.99	2.18
Pain or discomfort level 4	1.96	2.69	3.43	4.03
Pain or discomfort level 5	2.62	4.14	5.86	7.38
Anxiety or depression level 2	1.84	2.45	3.05	3.53
Anxiety or depression level 3	2.90	4.81	7.07	9.11
Anxiety or depression level 4	5.12	11.10	20.04	29.58
Anxiety or depression level 5	10.06	30.08	69.40	120.37
Observations	8021			
Log-Likelihood	9057			
AIC	18162			
Deviance	18114			
Correct predictive ability based on in sample prediction	51%			

Table 8.5 Final model for activities. Estimates are in odds ratio form

Consequence 2: relationships

The relationship variable is modelled as an ordinal variable. Using the ordered logit model, a chi square test comparison with the multinomial model shows that the proportional odds assumptions is violated (deviance of 137, with 34 degrees of freedom, p-value of <0.01). This is also shown graphically in Figure 8.6. If the proportional odds assumption was appropriate, the estimated log odds for each independent variable for each outcome level would be similar and thus the dots on each line would be close together (Harrell, 2001, p.336; Abreu et al., 2009). In the graph it is apparent that especially for the 'anxiety or depression' and 'pain or discomfort' domains

there are differences between the log odds ratios for different outcome levels. A stereotype model is estimated because the failure of the proportional odds assumption.

The results of the stereotype model are reported in Table 8.6 with the coefficients reported in log odds ratio form. The correlation between the probabilities of this model and the MNL model are high (between 0.84 and 0.99), which indicates that the stereotype model is appropriate and can be preferred due to being more parsimonious (Long and Freese, 2006, p.282).

The model contains a substantial number of inconsistent estimates, where worse health problems are associated with better relationship outcomes. Since the full model had some inconsistencies, the model was adjusted by combining the levels of inconsistent independent variables. For example, in the full model severe problems walking about has an odds ratio lower than one; which means that severe problems in walking about is associated with better relationship outcomes than no problems in walking about (although this odds ratio is insignificant at the 10% level).

Several models were fitted until consistent estimates remained. In the final model the levels of some, moderate, severe, and unable to/extreme are combined for the mobility, usual activities, and 'pain or discomfort' domains. For the self-care domain, the levels moderate, severe, and unable to are combined. The final model coefficients are reported in Table 8.7 and are presented as odds ratios for ease of interpretation.

The largest average odds ratio is for the anxiety or depression domain, followed by self-care, usual activities, and pain or discomfort, and mobility. The odds ratios for self-care, usual activities, and pain or discomfort, and mobility are similar and range from 1.01 to 1.19. The average 'anxiety or depression' odds ratio is 3.8. The size of the odds ratios indicate that compared to the base level of no problems, having problems in 'anxiety or depression' increases the relative odds of reporting worse personal relationships more than any other domain. The final model essentially indicates that

many of the health domains in the EQ-5D are not associated with personal relationships problems, when adjusting for an individual's 'anxiety or depression' level. For example, having any mobility problems is associated with very little increase in personal relationship problems.

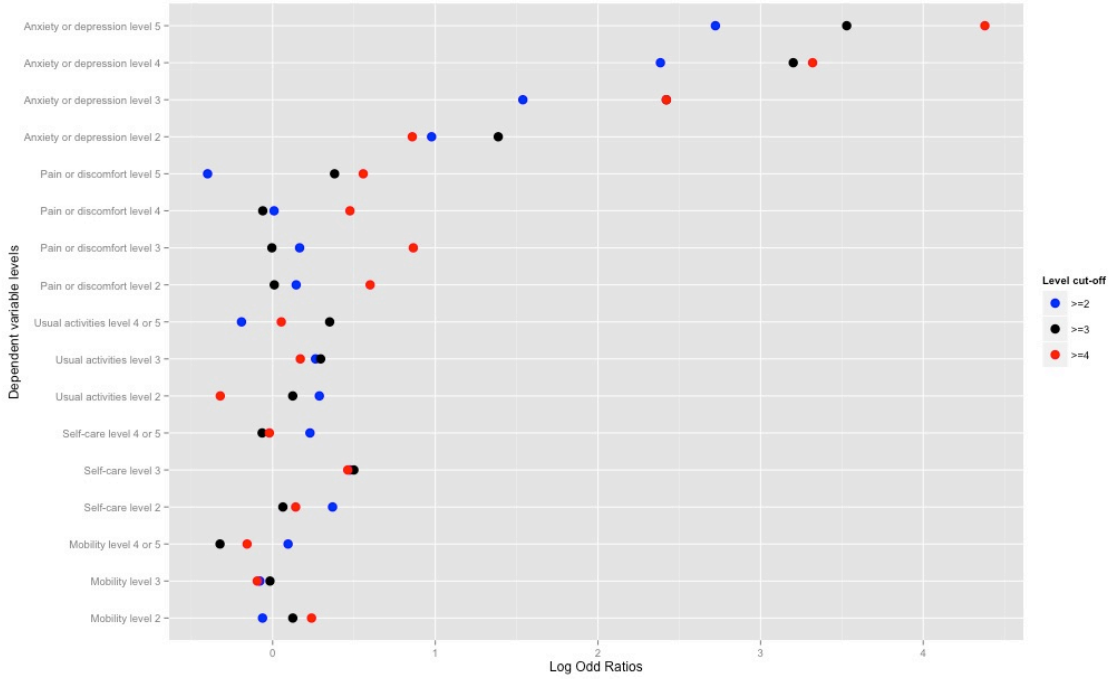


Figure 8.6 Proportional odds check for relationships

Explanatory variables	**Coefficient (log odds)**	**Standard error**	**P-value**
Mobility level 2	0.09	0.17	0.61
Mobility level 3	-0.11	0.24	0.66
Mobility level 4 or 5	-0.32	0.34	0.35
Self-care level 2	0.40	0.21	0.05*
Self-care level 3	1.00	0.30	0.00**
Self-care level 4 or 5	0.08	0.55	0.88
Usual activities level 2	0.40	0.16	0.01**
Usual activities level 3	0.62	0.25	0.01**
Usual activities level 4 or 5	0.34	0.36	0.35
Pain or discomfort level 2	0.27	0.15	0.08*
Pain or discomfort level 3	0.30	0.20	0.12
Pain or discomfort level 4	0.04	0.26	0.89
Pain or discomfort level 5	0.27	0.45	0.55
Anxiety or depression level 2	2.65	0.21	0.00**
Anxiety or depression level 3	4.41	0.30	0.00**
Anxiety or depression level 4	5.97	0.36	0.00**
Anxiety or depression level 5	7.06	0.46	0.00**
Theta 4	1.00		
Theta 3	0.65		
Theta 2	0.26		
Theta 1	0.00		
Constant 3	-6.69		
Constant 2	-2.70		
Constant 1	-0.61		
Observations	6845		
Log-Likelihood	6705		
AIC	13453		
Deviance	13409		
Correct predictive ability based on in sample prediction	53%		

* P-value < 0.1

** P-value < 0.05

Table 8.6 Full model for relationship. Reference category is the best outcome level

Explanatory variables	Odds of level 2	Odds of level 3	Odds of level 4
Mobility level 2 to 5	1.00	1.01	1.02
Self-care level 2	1.09	1.24	1.40
Self-care level 3 to 5	1.17	1.48	1.84
Usual activities level 2 to 5	1.11	1.30	1.49
Pain or discomfort level 2 to 5	1.07	1.19	1.31
Anxiety or depression level 2	1.99	5.60	14.21
Anxiety or depression level 3	3.13	17.48	82.08
Anxiety or depression level 4	4.60	45.64	360.16
Anxiety or depression level 5	6.11	93.31	1084.10
Observations	6845		
Log-Likelihood	6710		
AIC	13448		
Deviance	13420		
Correct predictive ability based on in sample prediction	52%		

Table 8.7 Final model for relationships. Estimates are in odds ratio form

Consequence 3: independence

As before, a stereotype model is fit because the proportional odds assumption is not appropriate (deviance of 158, with 34 degrees of freedom, p value of <0.01; see also Figure 8.7).

The full model is reported in Table 8.8. The correlation with the MNL is again high, ranging from 0.80 to 0.99. After adjusting for inconsistencies, the final model coefficients are reported in Table 8.9. In the final model, the levels of slight until extreme are combined for the pain or discomfort domain.

The highest average odds ratio for independence is the usual activities domain (4.57), followed by self-care (3.51), anxiety and depression (2.73), mobility (1.91) and pain or discomfort (1.22). The odds ratios indicate that reporting problems in usual activities problems is associated with reporting more problems in independence than other domains.

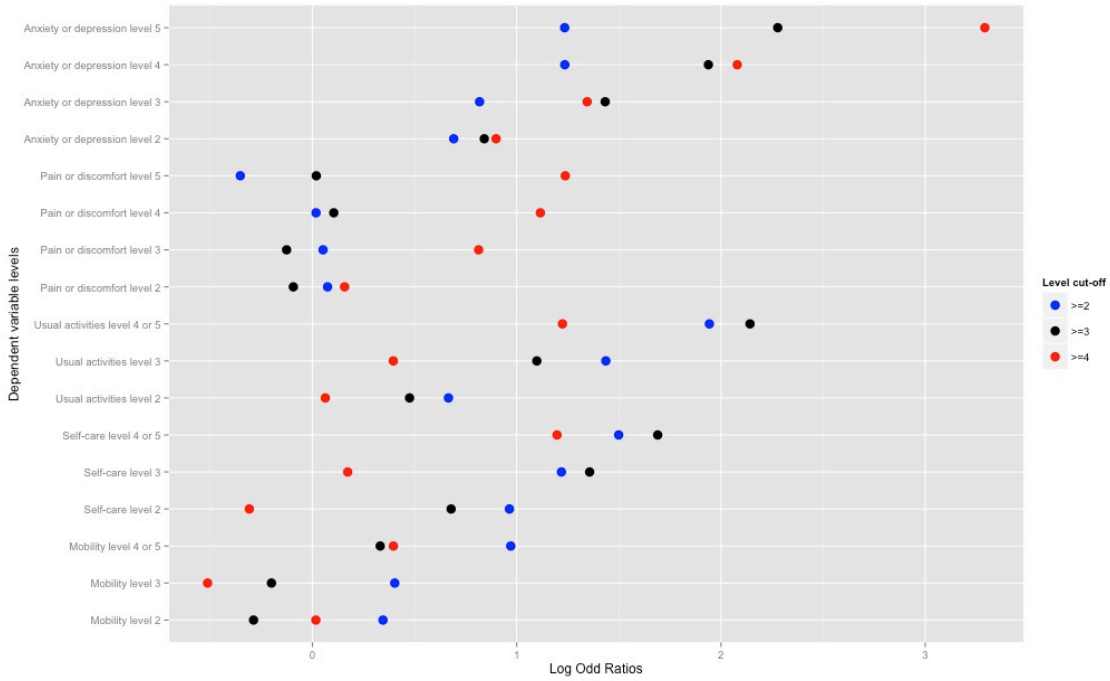


Figure 8.7 Proportional odds check for independence

Explanatory variables	**Coefficient (log odds)**	**Standard error**	**P-value**
Mobility level 2	0.44	0.14	0.00*
Mobility level 3	0.47	0.20	0.02*
Mobility level 4 or 5	1.45	0.33	0.00*
Self-care level 2	1.35	0.19	0.00*
Self-care level 3	2.42	0.33	0.00*
Self-care level 4 or 5	3.82	0.73	0.00*
Usual activities level 2	1.14	0.14	0.00*
Usual activities level 3	2.26	0.23	0.00*
Usual activities level 4 or 5	4.25	0.40	0.00*
Pain or discomfort level 2	0.11	0.12	0.35
Pain or discomfort level 3	0.07	0.16	0.67
Pain or discomfort level 4	0.18	0.22	0.41
Pain or discomfort level 5	0.02	0.47	0.97
Anxiety or depression level 2	1.25	0.12	0.00*
Anxiety or depression level 3	1.68	0.16	0.00*
Anxiety or depression level 4	2.61	0.25	0.00*
Anxiety or depression level 5	3.22	0.38	0.00*
Theta 4	1.00		
Theta 3	0.85		
Theta 2	0.50		
Theta 1	0.00		
Constant 3	-6.65		
Constant 2	-3.82		
Constant 1	-1.17		
Observations	6845		
Log-Likelihood	5637		
AIC	11317		
Deviance	11273		
Correct predictive ability based on in sample prediction	63%		

* P-value < 0.05

Table 8.8 Full model for independence. Reference category is the best outcome level

Explanatory variables	Odds of level 2	Odds of level 3	Odds of level 4
Mobility level 2	1.24	1.44	1.54
Mobility level 3	1.26	1.49	1.60
Mobility level 4 or 5	2.09	3.49	4.35
Self-care level 2	1.98	3.17	3.89
Self-care level 3	3.36	7.78	11.18
Self-care level 4 or 5	6.79	25.67	45.55
Usual activities level 2	1.77	2.63	3.12
Usual activities level 3	3.11	6.84	9.61
Usual activities level 4 or 5	8.46	37.22	70.54
Pain or discomfort level 2 to 5	1.06	1.10	1.11
Anxiety or depression level 2	1.87	2.88	3.47
Anxiety or depression level 3	2.32	4.16	5.36
Anxiety or depression level 4	3.72	9.27	13.75
Anxiety or depression level 5	5.02	15.39	24.96
Observations	6845		
Log-Likelihood	5637		
AIC	11312		
Deviance	11274		
Correct predictive ability based on in sample prediction	63%		

Table 8.9 Final model for independence. Estimates are in odds ratio form

Consequence 4: burden

The proportional odds assumption is rejected again (deviance of 220, with 51 Degrees of freedom, p value of <0.01; see also Figure 8.8).

The full stereotype logit model is reported in Table 8.10. The correlation with the MNL was good, ranging from 0.81 to 0.99. The full model had no inconsistencies and the final model coefficients are reported in Table 8.11.

The domains associated with the highest increased odds of reporting feeling like a burden is the 'anxiety or depression' domain (6.5), followed by usual activities (2.56), mobility (1.7), self-care (1.5) and pain or discomfort (1.4). Having problems in the 'anxiety or depression' increases the odds of reporting feeling like a burden the most from all EQ-5D domains.

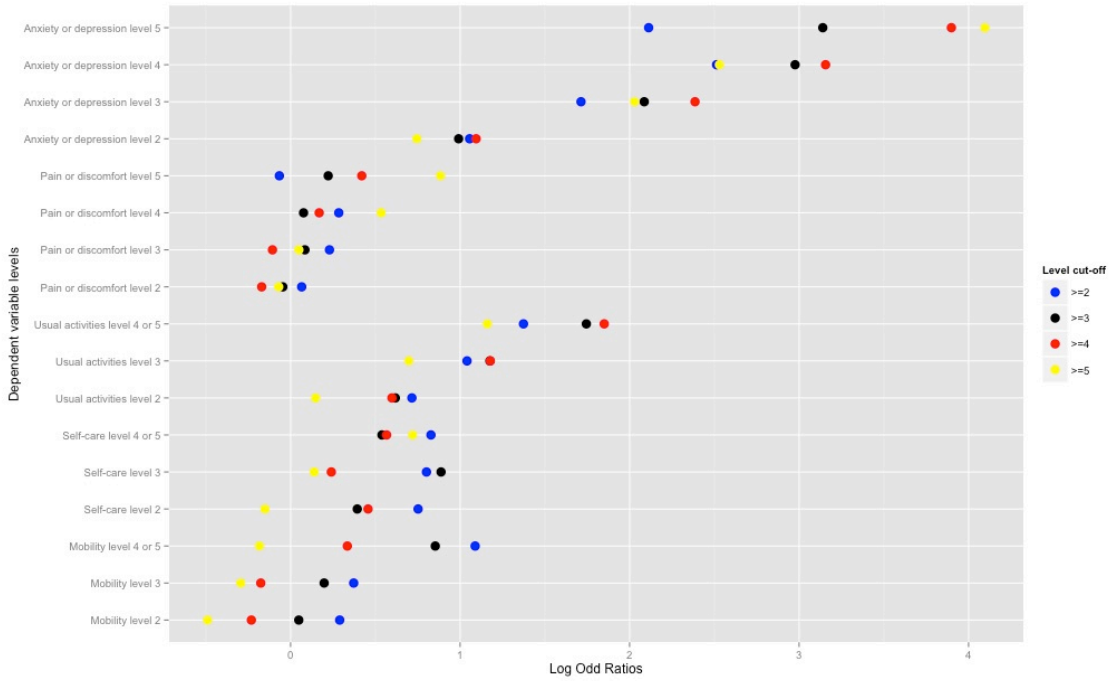


Figure 8.8 Proportional odds check for burden

Explanatory variables	**Coefficient (log odds)**	**Standard error**	**P-value**
Mobility level 2	0.36	0.15	0.01**
Mobility level 3	0.50	0.21	0.02**
Mobility level 4 or 5	1.71	0.32	0.00**
Self-care level 2	1.00	0.18	0.00**
Self-care level 3	1.27	0.31	0.00**
Self-care level 4 or 5	1.74	0.63	0.01**
Usual activities level 2	1.42	0.15	0.00**
Usual activities level 3	2.30	0.23	0.00**
Usual activities level 4 or 5	3.62	0.36	0.00**
Pain or discomfort level 2	0.10	0.13	0.46
Pain or discomfort level 3	0.34	0.17	0.04**
Pain or discomfort level 4	0.43	0.23	0.06*
Pain or discomfort level 5	0.65	0.46	0.16
Anxiety or depression level 2	2.16	0.14	0.00**
Anxiety or depression level 3	3.93	0.20	0.00**
Anxiety or depression level 4	5.88	0.28	0.00**
Anxiety or depression level 5	7.51	0.48	0.00**
Theta 5	1.00		
Theta 4	0.86		
Theta 3	0.65		
Theta 2	0.38		
Theta 1	0.00		
Constant 4	-7.69		
Constant 3	-5.43		
Constant 2	-3.43		
Constant 1	-1.38		
Observations	8022		
Log-Likelihood	7726		
AIC	14765		
Deviance	15451		
Correct predictive ability based on in sample prediction	58%		

* P-value < 0.1

** P-value < 0.05

Table 8.10 Full model for burden. Reference category is the best outcome level

Explanatory variables	Odds of level 2	Odds of level 3	Odds of level 4	Odds of level 5
Mobility level 2	1.15	1.26	1.36	1.43
Mobility level 3	1.21	1.38	1.53	1.65
Mobility level 4 or 5	1.92	3.02	4.34	5.52
Self-care level 2	1.47	1.91	2.37	2.73
Self-care level 3	1.63	2.28	2.99	3.58
Self-care level 4 or 5	1.94	3.09	4.47	5.71
Usual activities level 2	1.72	2.50	3.38	4.12
Usual activities level 3	2.40	4.42	7.21	9.95
Usual activities level 4 or 5	3.98	10.43	22.55	37.49
Pain or discomfort level 2	1.04	1.07	1.09	1.10
Pain or discomfort level 3	1.14	1.24	1.34	1.40
Pain or discomfort level 4	1.18	1.32	1.45	1.54
Pain or discomfort level 5	1.28	1.53	1.75	1.92
Anxiety or depression level 2	2.28	4.04	6.39	8.65
Anxiety or depression level 3	4.47	12.71	29.30	50.85
Anxiety or depression level 4	9.42	44.98	157.06	358.53
Anxiety or depression level 5	17.54	129.07	637.32	1828.47
Observations	8022			
Log-Likelihood	7726			
AIC	14765			
Deviance	15451			
Correct predictive ability based on in sample prediction	58%			

Table 8.11 Final model for burden. Estimates are in odds ratio form

Consequence 5: dignity

The proportional odds assumption is not rejected for the dignity variable (deviance of 11, with 10 Degrees of freedom, p value of 0.37). However, the graphical test reveals that there are some large differences between the log odds of the mobility domain in Figure 8.9. Therefore, and to maintain comparability with the other consequences, a stereotype model was fitted.

The full model did not contain inconsistencies (see Table 8.12). The model predictions show good correlation with the MNL model, ranging from 0.91 to 0.99. The final model is reported in Table 8.13 in odds ratio form.

The biggest influence on dignity is the mobility domain (7.9), with a similar odds ratio to 'anxiety or depression' (7.5). Pain or discomfort (2.5), usual activities (2.4), self-care (2.1) had smaller odds ratios.

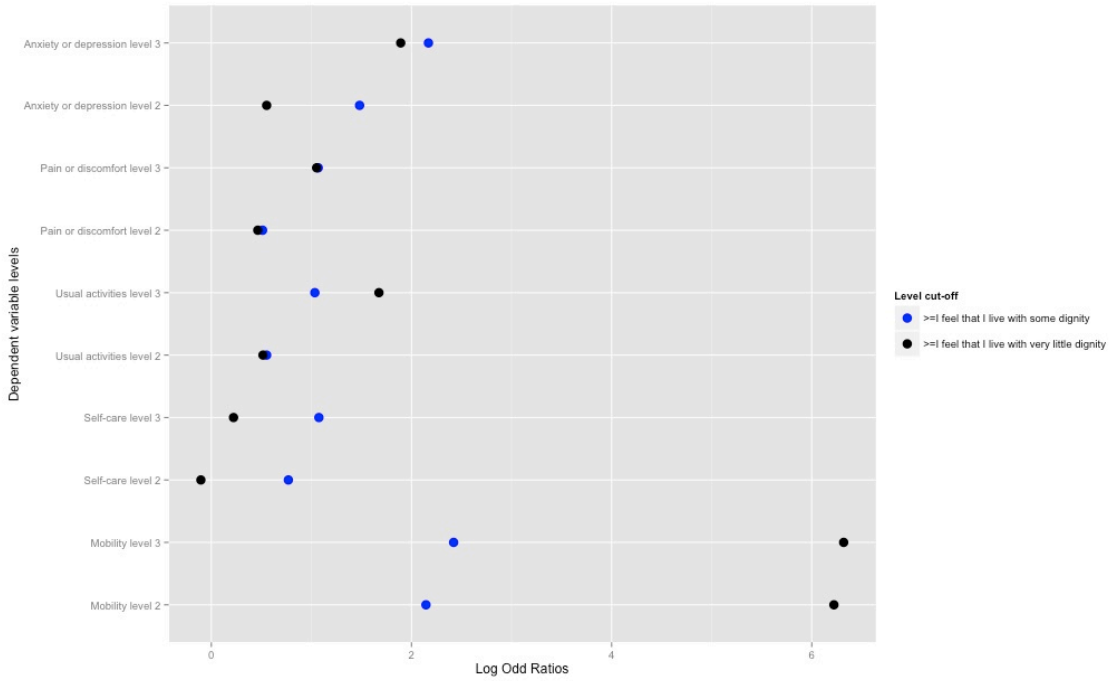


Figure 8.9 Proportional odds check for dignity

Explanatory variables	**Coefficient (log odds)**	**Standard error**	**P-value**
Mobility level 2	3.48	0.09	0**
Mobility level 3	3.79	2.21	0**
Self-care level 2	0.91	1.18	0**
Self-care level 3	1.42	1.86	0.06*
Usual activities level 2	0.88	1.90	0.05*
Usual activities level 3	1.85	0.55	0.1*
Pain or discomfort level 2	0.76	0.77	0.07*
Pain or discomfort level 3	1.59	0.87	0.31
Anxiety or depression level 2	1.93	0.93	0.05*
Anxiety or depression level 3	3.53	0.50	0.13
Theta 3	1.00	0.70	0.02**
Theta 2	0.60	0.49	0**
Theta 1	0.00	0.75	0**
Constant 2	-9.04		
Constant 1	-4.54		
Observations	276		
Log-Likelihood	228		
AIC	482		
Deviance	456		
Correct predictive ability based on in sample prediction	61%		

Table 8.12 Full model for dignity. Reference category is the best outcome level

Explanatory variables	**Odds of level 2**	**Odds of level 3**
Mobility level 2	8.00	32.45
Mobility level 3	9.63	44.25
Self-care level 2	1.73	2.50
Self-care level 3	2.34	4.14
Usual activities level 2	1.69	2.41
Usual activities level 3	3.01	6.33
Pain or discomfort level 2	1.57	2.13
Pain or discomfort level 3	2.59	4.93
Anxiety or depression level 2	3.17	6.88
Anxiety or depression level 3	8.24	34.06
Theta 3	1	
Theta 2	0.60	
Theta 1	0	
Constant 2	-9.0	
Constant 1	-4.54	
Observations	276	
Log-Likelihood	228	
AIC	482	
Deviance	456	
Correct predictive ability based on in sample prediction	61%	

Table 8.13 Final model for dignity. Estimates are in odds ratio form

Consequence 6: enjoyment

The enjoyment variable was modelled as a continuous variable and a beta regression was fitted. Enjoyment was rescaled from to a 0 to 1 scale, where one means 'completely happy' and zero means 'not at all happy'. Another transformation was applied because the beta regression assumes an open interval between 0 and 1, i.e. actual 0's and 1's are not expected. This transformation compressed the range of values to [0.0000623286,0.9999377].

Several inconsistent and insignificant coefficients were found in the full model (see Table 8.14) and those levels were combined. The final model is reported in Table 8.15. In this model the slight problems, moderate problems, severe problems, and unable to levels of the mobility and self-care domain are combined. For usual activities the levels

slight and moderate are combined. For 'pain or discomfort' the levels severe and extreme are combined.

Explanatory variables	**Coefficient**	**Standard error**	**P-value**
Constant	1.23	0.03	0.00**
Mobility level 2	0.04	0.04	0.32
Mobility level 3	0.06	0.06	0.32
Mobility level 4 or 5	0.19	0.09	0.04**
Self-care level 2	-0.17	0.05	0.00**
Self-care level 3	0.06	0.09	0.47
Self-care level 4 or 5	0.16	0.15	0.31
Usual activities level 2	-0.04	0.04	0.3
Usual activities level 3	-0.05	0.06	0.46
Usual activities level 4 or 5	-0.31	0.10	0.00**
Pain or discomfort level 2	-0.07	0.03	0.03**
Pain or discomfort level 3	-0.09	0.04	0.05**
Pain or discomfort level 4	-0.22	0.06	0.00**
Pain or discomfort level 5	-0.16	0.12	0.19
Anxiety or depression level 2	-0.78	0.03	0.00**
Anxiety or depression level 3	-1.46	0.04	0.00**
Anxiety or depression level 4	-2.22	0.07	0.00**
Anxiety or depression level 5	-2.85	0.09	0.00**
Observations	8008		

* P-value < 0.05

** P-value < 0.1

Table 8.14 Full model for enjoyment

Explanatory variables	Coefficient	Standard error	P-value
(Intercept)	1.23	0.03	0.00**
Mobility level 2 to 5	0.05	0.04	0.18*
Self-care level 2 to 5	-0.09	0.05	0.07*
Usual activities level 2 and 3	-0.05	0.04	0.2*
Usual activities level 4 and 5	-0.18	0.08	0.03**
Pain or discomfort level 2	-0.07	0.03	0.03**
Pain or discomfort level 3	-0.09	0.04	0.05**
Pain or discomfort level 4 or 5	-0.18	0.06	0.00**
Anxiety or depression level 2	-0.78	0.03	0.00**
Anxiety or depression level 3	-1.45	0.04	0.00**
Anxiety or depression level 4	-2.22	0.07	0.00**
Anxiety or depression level 5	-2.85	0.09	0.00**
Observations	8008		

* P-value < 0.05

** P-value < 0.1

Table 8.15 Final consistent model for enjoyment criterion

The model diagnostics, residuals compared to fitted values, are reported in Figure 8.10. There is some systematic relationship between predicted values and residuals, though substantially less than the OLS regression (OLS figures not shown). Where higher values are predicted the residuals are more negative and where lower values are predicted the residuals are positive. Figure 8.11 displays a normal quantile quantile plot of the residuals. Similar to the previous figure this suggests that the residuals may be slightly skewed.

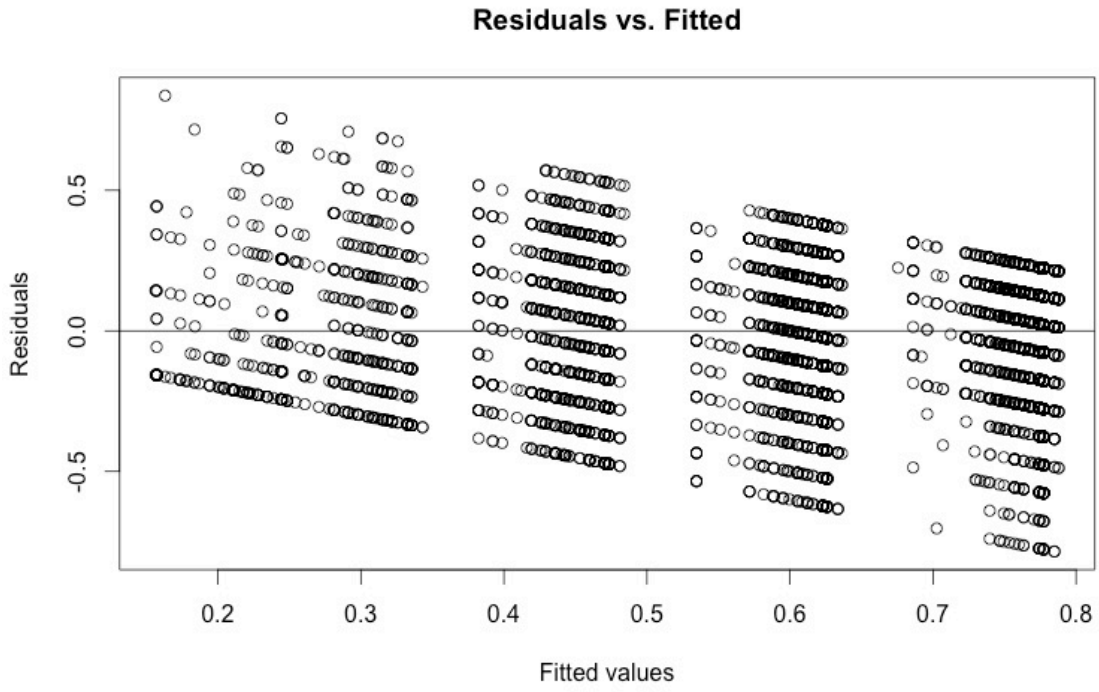


Figure 8.10 Plot of residuals compared to fitted values for enjoyment regression

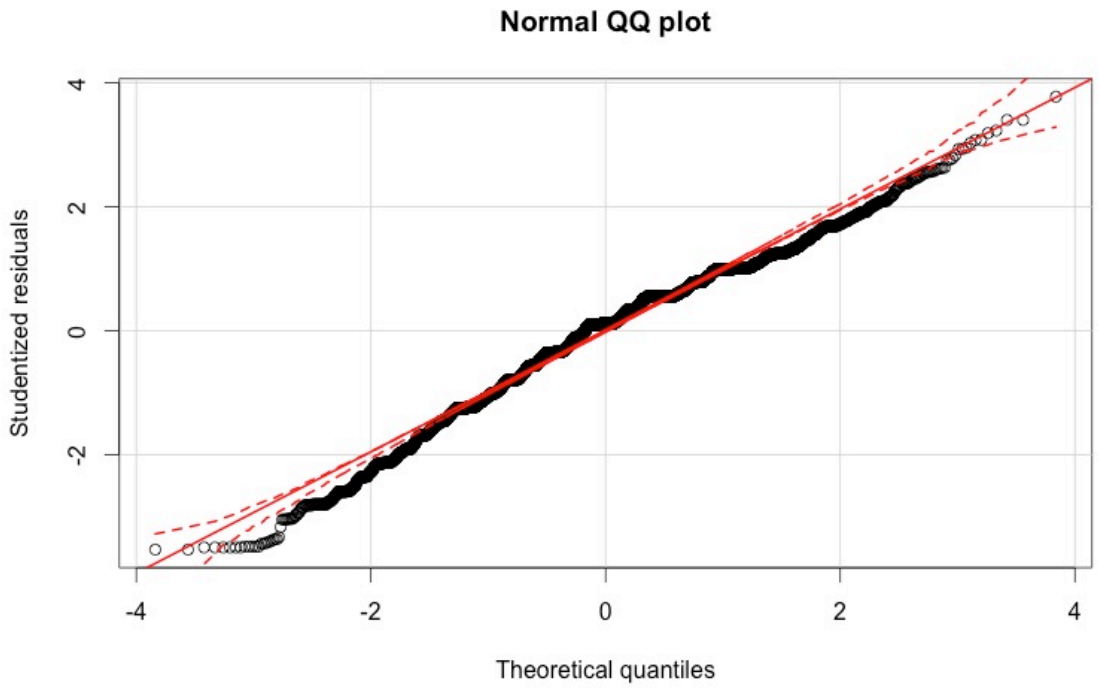


Figure 8.11 Quantile Quantile plot of enjoyment regression residuals. Line represents normal distribution

The interpretation of the coefficients is the change in the log odds of enjoyment by moving from the base level to the given EQ-5D level. The largest coefficient size for

enjoyment is the 'anxiety or depression' domain (average effect of -1.83), followed by usual activities (-0.11) and 'pain or discomfort' (-0.11) and self-care (-0.09). Mobility had a coefficient size of 0.05, meaning that mobility problems were associated with slightly better enjoyment outcomes. Comparing all domains 'anxiety or depression' is associated with the largest detriment in enjoyment, more than 10 times the next largest domain. The many inconsistent estimates in the full model indicate that many health problems are not associated with reductions in self-reported enjoyment. For example, increased problems in self-care beyond the slight level did not reduce enjoyment further.

A model for each consequence has now been estimated. These models will be used to predict the outcome of the six health states on each consequence.

8.2.3 Results of stage 2: comparing the six health states using experience and predictions

The six health states can now be ranked using the regression models based on the patient experience datasets. To calculate expected values for each health state, probabilities are estimated for each response level. Next, the probability of being in a level is multiplied by the number assigned to that response levels. and then these products are summed. The expected values for the dignity consequence had to be converted from EQ-5D-3L states to 5L states. The full probabilities results are reported in Appendix 19, Appendix 20, Appendix 21, Appendix 22, Appendix 23, Appendix 24.

The expected values were then ranked from best to worst. The rankings are reported in Table 8.16. The rankings provide a comparison of the reported experience of patients in the six health states for all six consequences.

For all six consequences, health states 31131 and 11331 are ranked by the patient experience as first and second, while health state 44545 is ranked as last. The rankings of the three health states 11334, 32322, and 44553 differ between the consequences.

For the consequences of relationships and enjoyment, health state 32322 is ranked above (3rd) health state 44553 (4th), which is ranked above health state 11334 (5th). For the consequences of activities and feeling like a burden, health state 32322 (3rd) is ranked ahead of 11334 (4th), which is ranked ahead of 44553 (5th). For the independence and dignity consequences health state 11334 (3rd) is ranked ahead of 32322 (4th) which is ranked ahead of 44553 (5th). There are thus similarities but also differences in the ranking of the six health states across the six consequences in the patient experience dataset.

The rankings for the expectations of the participants are reported in Table 8.17. The ranking is the same for all six consequences. The health states are ranked from best to worst as follows: 11331, 31131, 32322, 11334, 44553, and 44535. Health states 11331 and 31131 have similar scores, with average scores (SD) across all consequences of 78 (17) and 75 (18). Health state 32322 is ranked in the middle with an average score of 59 (22). Health states 11334, 44553 and 44535 are ranked lower, with average score of 31 (23), 14 (16) and 9 (23).

Health state	Relationships	Independence	Activities	Enjoyment	Burden	Dignity
31131	1	1	1	1	1	2
11331	2	2	2	2	2	1
32322	3	4	3	3	3	4
44553	4	5	5	4	5	5
11334	5	3	4	5	4	3
44535	6	6	6	6	6	6

Table 8.16 Rankings of states based on experience data

Health state	Ranking	Relationships	Independence	Activities	Enjoyment	Burden	Dignity
11331	1	81	78	71	73	81	86
31131	2	77	75	69	73	79	83
32322	3	62	60	54	54	60	65
11334	4	23	29	27	26	42	41
44553	5	18	12	12	12	16	17
44535	6	11	7	7	7	11	13

Table 8.17 Ranking and mean scores (on a 0 to 100 scale) of health states for each consequence based on expectation data. Ranking was the same on all consequence

8.2.4 Results of stage 3: comparing expectations and experience

The rankings estimated by the expectations of the participants and the rankings estimated using the experience of patients can now be compared. The two rankings are plotted in Figure 8.12. For all six consequences there is at least one difference in the ranking. For each pairwise comparison between two states, the correct or incorrect expectations of the participants are shown in Figure 8.13. Each criterion is discussed in turn below.

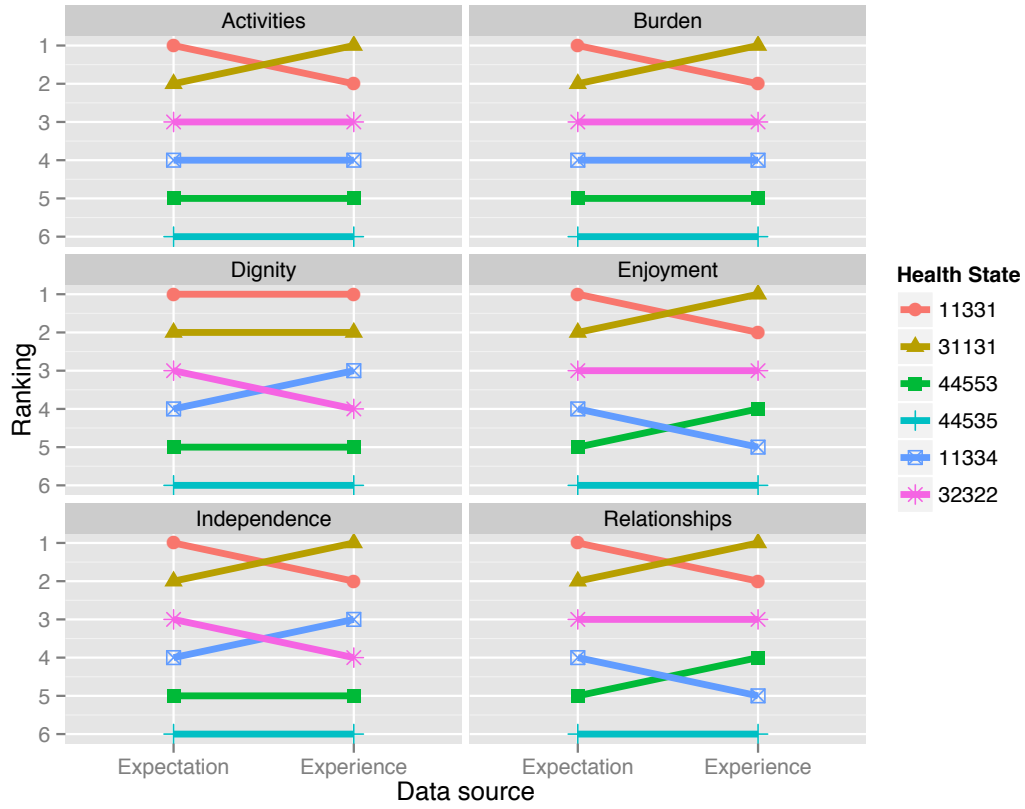


Figure 8.12 Comparison of expectations and experience rankings

Figure 8.13 Correct and incorrect expectations for pairwise comparison between states

Activities

In the activities consequence one difference was found between the two rankings. This was the difference between the ranking of states 11331 and 31131. Participants expected moderate mobility problems to be worse than moderate problems with usual activities but moderate problems in usual activities was associated with worse activities outcomes than moderate mobility problems.

Relationships

In the relationships consequence there are two differences in ranking between the participants' predictions and the patients' experience. Participants expected health state 31131 to be worse than 11331, while the experience data shows the opposite. Thus participants estimated moderate mobility problems to be worse than moderate problems with usual activities but usual activities problems were associated with worse outcomes than mobility problems. Secondly, participants estimated health state 44553 to be worse than 11334, but the patient experience of health state 11334 was associated

with worse outcomes. A one level detriment in 'anxiety or depression' is thus associated with worse personal relationships outcomes than the additional problems in the other four dimensions.

Independence

Two differences are found between the rankings of the participants and the patient experience data on the independence consequence. First, like the previous two consequences, there was the difference between states 11331 and 31131. Participants expected moderate mobility problems to be worse than moderate problems with usual activities but the reverse was true for independence. Second, participants estimated state 11334 to be worse than state 32322, but the health state 32322 is associated with worse independence outcomes according to the experience data. Therefore, participants overestimated problems in 'anxiety or depression' and 'pain or discomfort' compared to problems in self-care and mobility for the independence consequence.

Burden

Only one difference was found in the burden consequence. This was once again the difference between states 11331 and 31131. Participants expected moderate mobility problems to be worse than moderate problems with usual activities but usual activities problems were associated with being a bigger burden than mobility problems in the patient experience data.

Dignity

One difference was found for the dignity consequence. Participants estimated health state 11334 to be worse for dignity than health state 32322, but the reverse is true according to the experience data. Participants thus overestimated problems in 'anxiety or depression' and 'pain or discomfort' compared to problems in self-care and mobility for dignity.

Enjoyment

For enjoyment there are two differences in ranking between the participants' expectations and the patients' experience. Participants expected health state 31131 to be worse than 11331, while the experience data shows the opposite. Secondly, participants expected lower enjoyment in health state 44553 than 11334, but health state 11334 was associated with lower levels of enjoyment in the experience of patients. A one level detriment in 'anxiety or depression' is associated with worse enjoyment outcomes than the additional problems in the other four dimensions.

Overall, each consequence had at least one pairwise difference between the expectations of the participants and the experience of patients. This resulted in a total of nine differences. Many of these differences were repeated across the consequences, five of the differences were between health states 11331 and 31131, two were between 11334 and 44553, and two were between 11334 and 32322.

8.3 Discussion

8.3.1 Summary

This chapter reported on a study comparing the rankings of six health states on six consequences based on the expectations of participants and the experience of patients. A total of nine differences were found between the rankings, with one difference occurring five times. Given that there were six health states and six consequences, a maximum of 90 (5n or 15 per consequence * six consequences) pairwise differences could have been found. Therefore 10% of the total possible pairwise differences were found. This figure may be an understatement of the extent of differences between expectations and experience for two reasons. First, from the 15 potential pairwise differences per consequence, a total of eight pairwise comparisons involve a health state that dominates the other (i.e. in the pairwise comparison a health state has either equal

or less problems in every domain than the other health state, see Table 6.5 on page 150 for the list of dominated pairwise comparisons). The nine differences therefore represent about 21% of total non-dominated pairwise comparisons (9/((15-8)*6). Second, there are differences in severity between the non-dominated states and it is not likely that participants would misjudge the ordinal ranking between all non-dominated pairwise comparisons. Even 21% may therefore be an understatement of the true discrepancy between expectations and experience. Overall, this study indicates that although participants are not grossly misinformed about how health affects the six consequences, their expectations are not perfect even on an ordinal scale.

The most frequent difference in ranking was because participants underestimated the effects of moderate problems in usual activities compared to moderate problems in mobility. It is possible to speculate on whether participants would misjudge the other levels of these two domains. It would depend on participants' expectations of severity of the levels of each domain. There is evidence that the levels in the two domains are seen as equally severe. Herdman et al. (2011) asked respondents to score the EQ-5D-5L labels and domains on 0 to 100 scale. They note that the median values assigned to the levels of slight, moderate, severe, and unable to are similar for the mobility and usual activities domain. Second, it also depends on whether the experience of patients is worse for usual activities compared to mobility for each level of the two domains and the odds ratios of usual activities were generally higher than those of mobility. It is thus likely that the discrepancy between moderate problems with mobility and moderate problems with usual activities would apply to other levels of the two domains.

Both other differences in ranking involved health state 11334. For the consequences of enjoyment and relationships the effect of health state 11334 was underestimated compared to health state 44553. This meant that severe 'anxiety or depression' was underestimated compared to problems in the other four domains. But for the two

consequences of dignity and independence 11334 was overestimated compared to 32322. This mean that 'anxiety or depression' combined with 'pain or discomfort' was overestimated compared to problems in mobility and self-care for the consequences of enjoyment and relationships.

The regression results indicate that 'anxiety or depression' and usual activities are the two domains with the largest odds ratios for five of the consequences and mobility for one consequence. Comparing the odds ratios, it can be stated that for most consequences the relative odds of worse outcomes are most increased by having problems with 'anxiety or depression' and 'usual activities' compared to the other three domains. The results of the regression models have face validity given that some consequences would theoretically be expected to correlate with some health domains. For example, for the activities consequence it was the usual activities domain with the largest odds ratio and for enjoyment it was the 'anxiety or depression' domain with the largest odds ratio.

8.3.2 Comparison to literature

The only comparable literature is the literature comparing preferences to subjective well-being (i.e. enjoyment). The regression results of this chapter confirm the findings in the literature that there is a stronger association of the mental health domain (i.e. anxiety or depression) on subjective well-being compared to other health domains (Dolan and Metcalfe, 2012; Mukuria and Brazier, 2013; Cubi-Molla et al., 2014). Dolan and Metcalfe (2012) find that problems in 'anxiety or depression' are associated with a detriment in enjoyment about 10 times as much as mobility problems. Similarly, in this chapter the beta regression model estimated that odds of having worse enjoyment outcomes is higher for anxiety or depression problems than any other domain.

Dolan and Metcalfe (2012) argue that comparing values derived from preferences to measurements of SWB show that members of the general public undervalue mental health compared to physical health. In this chapter, participants underestimated the effect of having 'anxiety or depression' problems on enjoyment and personal relationships when comparing states 11334 to 44553. However, in this study anxiety or depression was not underestimated when comparing states 44535 to 44553 (i.e. comparing extreme pain or discomfort to extreme anxiety or depression directly). No literature was found on whether anxiety or depression is underestimated in other consequences. While it may be possible to say that participants underestimated the effect of anxiety or depression on enjoyment and relationships, it is not possible to state that they underestimate anxiety or depression on other consequences.

8.3.3 Implications for practice

The findings of this chapter indicate that participants are not perfectly informed about the health states they are valuing because their beliefs about the consequences of health states do not conform to the experience of patients. In particular, their beliefs about the consequences of problems in usual activities compared to problems in mobility are not in line with patients' experience. This can mean that interventions with usual activities improvement are undervalued compared to mobility improvements. This may not be problematic if improvements in the two domains are correlated. Similar problems exist for 'anxiety or depression' compared to physical domains. For consequences such as enjoyment and relationships participants underestimated problems with 'anxiety or depression' compared to problems in the physical domains. As a result interventions that improve mental health could be undervalued. These issues mean that the use of 'uninformed' preferences for valuation of the EQ-5D could result in sub-optimal policy recommendations in some circumstances.

8.3.4 Limitations of study

The limitations in this study include the wording of the questions for the consequences, the method of comparing the two datasets, and the study sample. As shown in Table 8.18 there are differences in the phrasing of the questions in the participant expectations and patient experience datasets. For example, the independence question of the ICECAP-A makes no mention of control but the experience question does. Some participants in the group meetings made a distinction between the two concepts. They felt that 'anxiety or depression' may not involve a loss of independence in the same way as mobility problems would, but that 'anxiety or depression' problems do involve a loss of a control.

Consequence	Expectations question	Experience questions
Enjoyment	Would you feel you are able to enjoy life?	Overall, how happy did you feel yesterday?
Activities	Would you feel able to do the things and activities that you want to do?	Cut down the amount of time you spent on work or other activities
Independence	Would you feel independent and in control of your life?	Being independent
Relationships	Would you feel satisfied with your personal relationships?	Love, friendship, and support
Dignity	Would you feel you can maintain your dignity?	Dignity
Burden	Would you be able to avoid being a burden on others?	How much of a burden do you feel you are to other people?

Table 8.18 Comparison of phrasing of questions in expectations and experience data for each criterion

Another difference between the expectations and experience is that the scales of the expectations and experience questions are different, and comparison could only be made on an ordinal basis. It may be that ordinal rankings are correct, while relative cardinal values are not.

The datasets used have limitations. The MIC datasets did not contain a large sample size in the worst response level of the EQ-5D-5L domains. The dignity dataset is fairly small, with a sample size of 276. Some EQ-5D-3L levels have few observations, for example only 16 (6%) of respondents had no mobility problems. Larger sample sizes, or more observations in certain cells, would make inferences more reliable. Both the MIC

and dignity datasets are cross-sectional. There is a potential for endogeneity and unobserved heterogeneity in this type of cross-sectional datasets (Wooldridge, 2009, pp.445-471). In addition, because of the cross-sectional data only association between the health and consequence can be assessed and not causality. For example, it is possible that the strong association between mobility problems and dignity are caused by some third factor and not the result of a causal influence of mobility problems on dignity. A longitudinal panel dataset would be useful to account for individual heterogeneity and can more easily assess causality (Wooldridge, 2009, pp.445-471).

There are limitations in the sampling of both the participants and secondary sources. Ideally, both samples should be similar and representative of the relevant populations. The sample of members of the general public in the expectations dataset is small and not representative of the general UK population. This makes it difficult to generalise the findings to the wider population. This is particularly important because average beliefs of members of the public are compared to average patient experiences. The MIC sample was an online sample. Although online-samples are being used more frequently in health economics studies (Mulhern et al., 2013), it may mean that self-selection is an issue. For example, the relationship between some health domains and enjoyment may be different for individuals who are more comfortable with technology and those individuals may be more likely to respond to an online survey.

The outcomes on the consequences were self-reported by the patients. This may be problematic if different groups use the response scales in different ways. This study assumed that self-reported outcomes amongst different groups can be directly compared.

Finally, the stereotype model assumes that one θ term can be used for all coefficients, meaning that the effect of the covariates compared to its baseline is assumed to be constant across all the outcome levels. This is not as flexible as the full

multinomial model but it does consider the ordinal ranking of the response levels of the consequences. The probability predictions compared to the multinomial model are sufficiently similar to prefer the more parsimonious stereotype model.

8.3.5 Future research

Future research can focus on addressing the limitations of this study to get a more accurate estimate of the extent that preferences are informed. This study has shown that useful results can be obtained using this design but the study had limitations. Most limitations can be overcome by further data collection. First, the sample could be improved by measuring expectations from a larger and more representative group of members of the general public. Patients could be recruited from a broader range of sources, rather than online recruiting as done in the MIC study. Ideally, a larger dataset could be used for the dignity question.

Second, a larger number of health states could be included. This study only included six states, which limited the number of comparisons that could be made. Adding more states also allows the use of more corner states, which would allow a direct comparison between domains. With the six states in this chapter only a direct comparison between the usual activities domain and mobility domain and 'pain or discomfort' and 'anxiety or depression' were available. A larger number of states can provide a better overall indication how participants are informed and whether there are certain domains about which they are less informed. Third, the study methods could be improved by using the same question and scales for both the expected and experience questions, which would allow for better comparison between the two. Findings from a larger study using similar methods would be more accurate and more generalizable.

This research could also be conducted with other generic preference-based measures such as the SF-6D and HUI. In particular the comparison between measures

within the skin and beyond the skin may be interesting. It may be that judging the consequences is difficult when using a within the skin type measures, such as the HUI.

Future research could also focus on developing more informed preferences by developing methods to inform participants of the consequences of health states. The findings indicate that it may be beneficial to provide members of the general public with more information about the experience of patients when valuing health states. In addition, it would be important to know how different informed and uninformed preferences are, and if those differences are of practical significance to cost-effectiveness analysis.

8.4 Conclusion

This chapter has developed and implemented a method to determine whether preferences over health states are informed. Expectations of a sample of members of general public were measured and compared to patients' experiences on six consequences. The expectations of participants did not always conform to the experience of patients, most consistently for the comparison between usual activities and mobility problems and between 'anxiety or depression' and other domains. This means that preferences are not entirely informed, and some health state values generated by choice-based methods may rest on false beliefs. Because choice-based methods rely on potentially uninformed preferences, the valuations generated by the choice-based methods may provide inappropriate recommendations in cost-effectiveness analysis.

This study focused on developing a new method and has limitations. The findings indicate that a fuller investigation with more health states is warranted to better assess the degree to which members of the public are informed.

Chapter 9

Discussion

9.1 Overview of themes discussed in this thesis

The aim of this thesis was to investigate three assumptions of the preference satisfaction theory of welfare in the context of choice-based methods for health. Three objectives needed to be achieved to answer the aim. The first was to describe how members of the general public value health states when using choice-based methods. The second was to assess whether there is a need for reflection and deliberation in health state valuation. The third objective was to develop and implement methods to determine whether preferences over health states are informed. The three objectives were achieved by using a sequential mixed methods study design. First, qualitative interviews were undertaken to understand how individuals value health states. The findings of the first study informed the design of the next two studies. The effect of reflection and deliberation on health state values was quantitatively tested in study two. A mixed method study in study two investigated the reasons for the quantitative findings. Finally, in study three expectations of members of the general public were compared to secondary data on the experience of patients.

In this discussion chapter the key findings of the previous four empirical chapters are summarised and reviewed in the context of existing research. The key contributions of this thesis to knowledge and the methodological contributions of this thesis are highlighted. Next, the implications of the findings for using preferences to value health are discussed. Finally, the limitations of this thesis are reviewed and recommendations for further research are made.

9.2 Key findings and contributions to existing knowledge

The following three sections describe the key findings for each study of this thesis. The findings are compared to existing literature and the contributions of the findings to existing knowledge are highlighted. Following this is a section on methodological contributions.

9.2.1 Study 1: how do members of the public value health?

This study was intended to investigate the range of underlying beliefs and factors that people consider when valuing health, the relationships between those beliefs and factors, and how those beliefs affect people's preferences. Understanding the thought processes of individuals valuing health was motivated by an implicit assumption in some of the economics literature that preferences are matters of taste that are not subject to discussion or criticism. The thought processes of members of the general public while valuing health states was investigated through a set of 21 semi-structured qualitative interviews with a think-aloud protocol.

Preferences over health states were not matters of taste and participants were found to consider a range of factors and beliefs when valuing health, many of which were not health related. These factors ranged from family circumstances to the effect of health states on enjoyment of life. There was a relationship between some of these factors, which resulted in grouping some factors as 'conversion factors' and some as 'consequences'. Conversion factors are factors in a participant's personal and social situation that determine how ill health affects their life. Consequences are the effect of ill health on an individual's life. The six of the most frequently mentioned consequences were activities, enjoyment, independence, relationships, dignity, and avoiding being a burden.

An explanatory framework was developed to explain the process by which beliefs of participants influence the value of health states. The explanatory framework contained several stages, from the act of participants reading the health state until they expressed their preferences. Participants were observed to interpret and concretise a health state by combining the health state description, their past experience of ill health, and their imagination. The interpretation was then combined with a participant's beliefs about conversion factors. Conversion factors were grouped under personal interests and circumstances, other people's reaction to ill health, adaptation, and available support. By combining the interpretation and conversion factors participants were able to determine the consequences for each health state and these consequences were then weighed. Ultimately, participants' health state preferences were constructed based on their beliefs about how health would affect their life. Health state values may be best described as "the desirability of life given a particular state of health" (Tsevat, 2000).

The key contribution to knowledge in this study was to highlight the relationship between the various non-health factors and beliefs and people's preferences over health states. Previous think aloud have indicated that participants consider a wide range of factors when valuing health but a key contribution of this study was the development of an explanatory framework. The framework distinguishes two types of non-health factors, conversion factors and consequences. The role of conversion factors and consequences are different in valuing health. Beliefs about conversion factors affect the consequences of a health state and the consequences of a health state determine how good or bad health states are. The role of non-health factors in valuing health has thus been explored in greater depth than previously explored in the literature.

The findings of this study provided inputs for both study two and three by showing how preferences could be discussed and questioned. First, the explanatory framework emphasised that health state valuation is a cognitively complex and subjective task, and

reflection and deliberation, including MCDA, may be guide participants in constructing their preferences. The importance of the consequences in valuing health suggests that reflection and deliberation should focus on the consequences. Six consequences were most frequently mentioned and this was taken as an indication of their importance. It was decided that the six were used as 'criteria' in a MCDA exercise in study two. For study three, the findings of this study suggested having informed preferences can be conceptualised as being informed about the effect of ill health on the six consequences.

9.2.2 Study 2: what should be the role of reflection and deliberation in health state valuation?

Motivated by the literature on preference construction, the objective of the second study was to investigate whether there is a role of reflection and deliberation in health state valuation. To reach this objective, a reflection and deliberation exercise was designed and its effect on health state values tested. Then a mixed method study investigated the reasoning behind the quantitative findings. Together the findings of this study provide evidence for whether choice-based methods should incorporate reflection and deliberation.

The reflection and deliberation exercise was conducted in a group setting. Participants attended one of a series of group meetings. Each group meeting consisted of four stages: (i) participants individually valued health states, (ii) participants individually completed the MCDA booklet as a reflection exercise, (iii) participants deliberated on their answers in the MCDA booklet as a group, and (iv) participants individually valued the same health states again.

MCDA was chosen as the method to encourage reflection and guide deliberation. MCDA was chosen because it serves as a useful tool to guide participants from their potentially more general values to their more articulate values. The MCDA task asked

participants to score health state on six criteria, with criteria being the consequences that participants had mentioned in study one (activities, enjoyment, independence, relationships, dignity, and avoiding being a burden). Participants were also asked to weigh the six criteria. This exercise encouraged participants to reflect on how the health state would affect their life. The MCDA exercise also served as a tool to guide the discussion during the deliberation of the group meetings.

Reflection and deliberation did not result in statistically significant changes to health state values at the aggregate level. Mean TTO values before and after the group meeting were compared and the changes ranged from -0.058 to 0.04 and were below 'small' using effect size guidelines by Cohen (1988). This range of changes is likely to be of no practical significance to most economic evaluations in most circumstances, although it is difficult to judge what level of change in health state values would change cost per QALY calculations sufficiently to change funding decisions. At the individual level a large number of changes to health state values were found. The findings suggest that individual level values may be too unstable for use in individual decision-making but that aggregate values are stable enough to be used in cost-effectiveness analysis.

The key contribution of the quantitative part of this study was to extend knowledge of the effect of reflection and deliberation implemented as a group-based exercise to a more widely used health description system (i.e. EQ-5D), valuation method (i.e. TTO), and aggregation method (i.e. mean values). In the literature review it was found that from the group-based exercises four papers found aggregate level changes to health state values while one did not. The four papers that found changes were all unconventional in some way, Krabbe et al. (1996) used a voting system, Akunne et al. (2006) required consensus, McIntosh et al. (2007) used the CLAMES classification system, and Robinson and Bryan (2013) used the PTO. Perhaps the contradictory results in the literature can be explained by those factors. Regardless of the explanation,

this study provides some evidence that those results cannot necessarily be generalised to other descriptive systems and valuation methods, including those using the TTO and EQ-5D.

This study also used a mixed methods approach to explain the quantitative findings. Qualitative analysis of the audio recordings of the discussion period of each group meeting was combined with the quantitative pre and post group meeting TTO values. The qualitative analysis identified several themes that could explain the quantitative results. It was argued that the lack of large systematic changes in health state values were because disagreements amongst participants were not solved but either ignored or participants 'agreed to disagree'. Health state preferences were seen as personal and participants did not hear much new surprising information. The non-systematic changes at the individual level were explained by the fact that participants expressed uncertainty about their valuations and knowledge, partly due to their lack of experience. The extent of positive or negative discussion around health states predicted change on the ranking of one pair of health states but not another. Overall, it was concluded that large changes are not expected after reflection and deliberation because health state valuation is considered personal and that the discussion is centred on an exchange of subjective attitudes and not an exchange of factual knowledge or beliefs.

The key contribution of the mixed methods study is to explain the reason why reflection and deliberation did not change values by understanding the nature of reflection and deliberation that took place. In particular it suggests that preferences are considered personal and that there is little factual knowledge exchange. This study strengthens the findings from two previous qualitative studies on reflection and deliberation. Stein et al. (2006) comment that in their study participants shared little new information to the rest of the group and that most participants valued the group deliberation for procedural reasons. Robinson and Bryan (2013), in a study using the

PTO, conclude that participants were not likely to change their valuations because of other participants' valuations or influence of other group members but that discussing the implications of the participants' valuations for treating certain groups of patients instead of others did prompt changes. In this study it was found that participants do not seem to be influenced by other group members but because of the use of TTO there were no direct equity implications of treatment decisions to consider.

The overall findings of this study suggest that reflection and deliberation do not appear to invalidate the assumption of completeness because aggregate values do not change compared to one-off interview methods that do not include reflection and deliberation. Although people are uncertain about their values, the process of deliberation does not necessarily help them because they consider their preferences personal and do not seem to accept other people's experiences.

9.2.3 Study 3: Informed preferences

The final objective of this thesis was to assess whether health states preferences of members of the general public are informed. This was motivated by the assumption of rational preferences may not be sufficient in welfare economics. Because there was no method of testing whether preferences are informed found in the literature, a method had to be first developed.

First it was argued that the findings of study one provide a test of having informed preferences, namely being informed about the consequences of health states. To determine whether members of the general public are informed about the consequences of health states their expectations were compared to the experience of patients. Participants ranked six health states on six consequences and this was compared to rankings based on the experience of patients. The six consequences were the same six consequences as in study one (activities, enjoyment, independence, relationships,

dignity, and avoiding being a burden). Participants' scores in the MCDA booklet collected in study two were used as data for participants' expectations. The rankings of the experience of patients were calculated using self-reported data from secondary sources by applying stereotypy logistic and beta regression models. The rankings of participant expectations and patient experience were compared for each consequence to determine whether participants were informed.

The findings suggest that participants were not perfectly informed about the health states. Each health state was compared in a pairwise fashion with other health states to see if the pairwise ranking conformed to the patient experience data. It was found that in about 21% of non-dominated pairwise comparisons the expectations of participants did not conform to the experience of patients. In particular, participants undervalued problems in the 'usual activities' domain of the EQ-5D-5L compared to problems in the 'mobility' domain of the EQ-5D-5L. For example, moderate problems in usual activities are associated with a lower enjoyment than moderate problems in mobility but participants in the study thought the reverse was true. Participant also made pairwise errors when comparing problems in 'anxiety or depression' to several other domains. Participants could not be considered completely informed about the consequences of the health states, even at an ordinal level.

The key contribution to knowledge of this study of the thesis was to show that preferences over health states cannot be considered entirely informed. The findings cannot be directly compared to existing literature because no literature was found. There are similar studies comparing valuations based on preferences to valuations based on subjective well-being. This literature has consistently found that 'anxiety or depression' has a bigger effect on enjoyment than is estimated using participants' preferences. In this study it was also found that participants underestimated the effect of anxiety or depression on enjoyment and relationships compared to increased

problems in several physical domains. Participants did not underestimate the effect of anxiety or depression for all consequences. The findings of this study thus suggest that people underestimate the consequences of anxiety or depression for enjoyment and relationships but not necessarily for all consequences.

9.2.4 Methodological contributions of this thesis

To accomplish the aims of this thesis, a set of methodological contributions were also required. Novel contributions in the thesis are highlighted below.

The use of mixed methods. This thesis has shown the benefits of a mixed method design for answering research questions that cannot be answered using either quantitative or qualitative approaches alone. The use of qualitative methods in study one was used to design both study two and three. The qualitative data highlighted the aspects of health state valuation that may require reflection and deliberation. This meant that the group meeting could be more formally structured than was the case in the existing literature. Study one also highlighted which aspects of health states people should be informed about. Study two contained a concurrent mixed methods study that integrated qualitative and quantitative data on the same cases. To our knowledge, this was only the third application of mixed methods where the two components were integrated at the interpretation stage in health economics (Robinson et al., 1997; Ryan et al., 2009). Applying a mixed methods approach using the same cases allowed a better understanding of what happened during the group meetings and allowed the researcher to draw more accurate implications than from the quantitative results alone. This mixed methods study is novel because of the use of qualitative data from groups and because the systematic use of the triangulation protocol (O'Cathain et al., 2010). The previous two mixed methods studies in health economics used individual interviews and not group meetings. The use of group data involves additional challenges, such as the

design of the study and the mode of data collection that may have to more structured than in interviews where more flexibility can be expected, decisions over aggregating the data across groups, and analysis of individual or dialogues. The use of the triangulation protocol had not been previously systematically implemented because in Ryan et al. (2009) integration was achieved using a mixed methods matrix (O'Cathain et al., 2010), where data is integrated at the analysis stage. The study by Robinson et al. (1997) could be referred to as using the triangulation protocol, but this was not implemented in a systematic fashion. For example, the quantitative and qualitative data were not separately analysed and then systematically compared. This study implemented the triangulation protocol by connecting a set of qualitative themes to the quantitative data using hypotheses. The stages followed in the analysis reported in Chapter 7 (analysis of qualitative data, summary of themes, hypotheses generation, and testing hypotheses) can be applied in a variety of other contexts and may be useful to future studies. Explicit qualitative work can prevent the need for speculation by researcher for explanation of the findings (Coast, 1999). Further use of mixed methods to answer research question in health economics is feasible and can be valuable.

The use of qualitative methods for providing an explanatory account.

Previous qualitative studies had been conducted to understand how people value health. Most of those focused on providing a descriptive overview of the themes found in the data. This thesis went further by attempting to assess linkages between the themes rather than only providing a description of the data. For example, it was argued that the EQ-5D health state together with previous experience and imagination are used to develop a concrete version of the health state. In addition, three techniques (looking for outliers, looking for negative evidence, and finding and testing rival explanation) were used to critically review the explanatory account in an attempt to balance the added interpretation needed in explanatory account. The findings demonstrate the value of

qualitative work in health economics and how qualitative methods can be further extended.

The development of a reflection and deliberation exercise, including the use of MCDA. A reflection and deliberation group exercise was designed in study two, which included the use of MCDA. The MCDA exercise had not been applied in health state valuation before as a method of encouraging reflection. Although MAUT methods can be considered a type of MCDA an important difference between using MAUT to value health and the MCDA implemented in this thesis is that the health states were not valued directly, rather the health states were scored on their consequences. The MCDA booklet was also used as a tool to guide deliberation. The MCDA exercise was found feasible because all participants completed the task. The task could be used and further developed in a larger study.

The development of a method to assess whether preferences are informed. The combination of scores elicited during the MCDA exercise and secondary data was used to assess whether preferences are informed. The literature review showed that no such method has been previously implemented. The method was found to be feasible and can be further developed and implemented in a larger study.

9.3 Implications of findings for methods of health state valuation

The implications of this thesis for conventional choice-based methods of valuing health are mixed. The assumption that preferences are complete was not found to be violated. Although preferences were not found to be entirely complete at the individual level, reflection and deliberation did not change health state values and therefore conventional methods can be said to be valid in that respect.

The implications of the failure of having informed preferences are arguably the most important for the validity of choice-based methods. The findings of study three

suggest that choice-based methods used to value health may be based on preferences that cannot be considered to be fully informed. As a result of using partly uninformed preferences over health states certain interventions may be over- or undervalued. When measuring benefits using the EQ-5D-5L, improvements in usual activities may be undervalued compared to improvements in mobility and 'anxiety or depression' may be undervalued compared to physical domains. The extent that not being fully informed causes a problem for cost-effectiveness analysis is hard to judge because this study cannot estimate how different informed preferences would be and what effect that would have on cost-effectiveness figures. Therefore, one implication is that further research should be conducted to assess the importance of using preferences that are not entirely informed.

The implication of the first study is that health economists should spend more time directly investigating preferences because preferences are not simply matters of taste. Indeed, the design of studies two and three were based on study one. Another implication of study one is to consider alternative descriptive systems. One suggestion in the literature is the use of subjective well-being measures (Dolan and Kahneman, 2007). The finding of study one would suggest that subjective well-being may not be broad enough because participants care about more than subjective well-being (independence, avoiding being a burden, relationships, activities, and dignity). A measures such as the ICECAP (Al-Janabi et al., 2012) describe a wider variety of important consequences of health and have similarities to the six consequences of study one and may be more appropriate. However, a disadvantage would be the possible lack of responsiveness in such a measure when dealing with health interventions (Mukuria et al., 2015). It may be that the ability of members of the general public to adequately assess the non-health consequences of health has to be traded off with the

responsiveness of a health-related questionnaire to determine whether a well-being type or health measure is preferred.

9.4 Limitations

The key limitations of this thesis are the sample used and the design of the studies. For both study two and three it would be preferable to use a larger sample and one that is more representative of the general population. The samples for study two and three are relatively small for a quantitative study, though they are similar to some previous studies of the same nature (Robinson and Bryan, 2013). A larger sample would result in a more power, which would be able to detect statistically significant changes of smaller magnitudes.

The sampling method should be extended to use random probability sampling to achieve a representative sample. Compared to the UK population (Office for National Statistics, 2011) the sample in study two and three contains more females, those with degrees, students, and retirees, but fewer employed people. Compared to the UK population (Office for National Statistics, 2011) those aged 60 and above were oversampled, while those aged 30 to 59 were under sampled. A representative sample allows the findings to be generalised to the UK population. That is important because their preferences are used to value health states by reimbursement agencies such as NICE (NICE, 2013a). A more representative sample is also important because it is possible that certain segments of society have more incomplete preferences and benefit more of reflection and deliberation. Similarly it is possible that certain segments of society are more informed about health states than others. Therefore to understand whether the public as a whole needs reflection or deliberation or is informed requires a representative sample. It is possible, for example, that more educated than average individuals are more informed and this may bias the degree to which they appeared

informed about health states. Random sampling also reduces the possibility of bias in the findings that can result from self-selection bias, where individuals volunteer themselves for a study. Using a more representative sample is needed to generalise the findings to the general public in the UK. The sample could be recruited using a postal mail method or by randomly selecting individuals from the local registry of electorates. Although resource constraints caused the limitations in the sample they could be removed in a further study.

There were limitations in the patient experience datasets used. The sample is not representative of the population. In particular the use of an online recruiting method based on individuals who have registered for a panel can have significant self-selection bias. For example, those who are severely ill may not answer questionnaires online. The full dataset was used, which involved individuals in several countries (Australia, UK, USA, Canada, Norway, and Germany). The use of the full dataset increased the sample size, but included non-UK residents. The dignity dataset was smaller and was also unrepresentative, especially because the data was collected from patients with pressure ulcers. There may be the possibility that the results found using this sample are not necessarily generalizable to the UK population. The significance of this depends on whether the relationship between health and the consequences of ill health are believed to differ amongst sub-groups, but there is too little evidence to answer this. In this study the health states were ranked on an ordinal level, thus any cardinal differences resulting from a different sample would not be a problem for this study.

The patient experience datasets were cross-sectional, which limits the ability to assess causality. Ideally the data would be from a longitudinal panel data (i.e. track the same people over time) because that would give researchers the ability to assess endogeneity and control for unobserved time-invariant variables, which increases the likelihood that the findings are causal and not just correlational (Wooldridge, 2009,

pp.445-471). There may also be a quality concern because of the use of online data collection, but the data quality has been reported to be of good quality, perhaps due to the stringent quality control data processing that was applied (Richardson et al., 2012, pp.24-25; Richardson et al., 2015).

Resource constraints also limited the types of designs that were possible in study two and three. There is the potential to design the reflection and deliberation exercise differently and allow for more reflection and deliberation. For example multiple meetings with the same participants could be conducted or MCDA valuations could be calculated and shown to the participants. It is possible that those designs would be more helpful to participants, although the evidence from the mixed methods aspect of study two would not necessarily support this.

It would also be possible to change the study design by including a control group. The use of a control group would be useful because other research has shown that individuals' preference change to cohere with their choice (Holyoak and Simon, 1999; Simon et al., 2004). For example, Simon et al. (2004) found that after participants are asked to choose between two job offers, the attributes which were most attractive for the chosen job offer were considered more important. Extending this to the EQ-5D, it could be that once a participant has chosen to value health state 11331 greater than health state 31131, subsequently they come to find the attribute of usual activities more important than mobility, more so than before they had made that choice. This means that allowing people to value health states and then to reflect and deliberate on them may bias results towards finding no change (although the extent of changes at the individual level suggests that this may not have been entirely the case in this thesis). A more reliable study design would be the Solomon four-group design (Dimitrov and Rumrill, 2003), which involves three control groups. The three control groups can control for various confounders, such as the possibility that time between the two TTO

tests or the pre-testing affect the results. Using a control group increases the certainty that the findings are the pure effect of reflection and deliberation.

Lastly, the design of study three could be improved. Data collected for study three was nested in study two and therefore the scope for the design of the task was limited. For example, the use of only six states limited the number of pairwise comparisons that could be made. Because different scales were used for the experience and expectations data, the comparison had to be restricted to ordinal rankings. These two limitations meant that the overall view on how informed people are is limited. This study should be seen as an exploratory study and a larger study could be designed with more states.

9.5 Further research

Many of the methods used in this PhD were exploratory and therefore one set of future research recommended is to expand on the methods. In particular, the design of the testing for informed preferences could be done in a more systematic way. This would involve the use of a larger and randomly selected sample, more health states, and the use of the same scale on both experience and expectations. This would provide a broader understanding of the nature and extent of the degree to which preferences of members of the general public are informed.

New avenues for research are suggested by the findings of this thesis. First, the framework in study one could be tested quantitatively. For example, some of the non-health consequences or conversion factors could be specified along with the health state to see whether the health state values will change. This could be similar to recent work by Dolan et al. (2013) that tested how specifying life satisfaction in a health state affected health state values. Testing the framework quantitatively can provide more evidence for its validity.

Second, the generalizability of the six consequences can be investigated. The six consequences were important for both study two and three of this thesis. If further research using these methods is to take place, it is important to understand the universality of the six consequences. The universality could be established by completing a similar qualitative study as in study one with a larger and more representative sample. It could also be established by survey methods, for example people could be asked to value health states and then write in their top six considerations.

Third, methods for obtaining more informed preferences could be tested. One avenue to achieve more informed preferences is to give participants information on the six important consequences of ill health before valuing health. Information could, for example, be provided in a similar fashion as done by McTaggart-Cowan et al. (2011) who provided participants with information from patients by using video clips of patients (although that study focused only on adaptation and not broader consequences of health). Reflection and deliberation were studied in this thesis in the context of obtaining complete and articulated preferences. However, there have been also suggestions to use reflection and deliberation as methods of developing informed preferences. The findings from study two suggest that these methods of reflection and deliberation are not likely to be the best method to achieve more informed preferences. The changes in health state values are small and the qualitative work indicates that the deliberation tends to focus on exchange of personal views rather than knowledge. The implication is that if preferences are to be better informed, alternative methods should be considered. This line of research could also investigate how different informed preferences are to preferences elicited using conventional methods that do not provide additional information to participants.

Finally, the methods in this thesis could also be applied to outcome measures other than the EQ-5D. Given the reliance of NICE on EQ-5D it was chosen as the outcome measure to be used. However, other questionnaires are used in cost-effectiveness studies. It would be important to see if the findings in this thesis can be generalised to other established outcome measures such as the SF-6D, the HUI, or the AQoL. The method in which health is described may affect the difficulty for people to judge the impact of health on their life. Similarly, different methods for valuing health such as DCE or SG could be investigated.

Overall the following areas of further research are recommended:

- Test the degree to which health state preferences are informed in a larger study with an improved study design
- Quantitatively test the explanatory framework of study one
- Investigate whether the six consequences are generalizable to a representative sample of the general population in the UK or other countries
- Test whether informed preferences are different from manifest preferences by providing information on how the six consequences affect health
- Investigate the extent to which 'uninformed' and informed preferences are different to determine the practical significance of using uninformed' preferences to provide policy recommendations
- Apply the methods in this thesis to other outcome measures and valuation methods used in health economics

9.6 Conclusion

The validity of choice-based methods for valuing health depends on the validity of its underlying assumptions. Three assumptions were tested in this thesis. It was found that

preferences over health states are not purely matters of tastes but depend on beliefs about conversion factors and consequences of health states. An indirect test of the assumption of complete preferences over health states at the aggregate level did not invalidate the assumption because reflection and deliberation did not result in statistically significant changes to health state values. Although the previous two findings do not suggest any threats to the validity of choice-based methods for valuing health, it was found that there is some evidence that preferences are not entirely informed and thus preferences cannot be assumed to be informed. Therefore, conventional choice-based methods may recommend sub-optimal resource allocation choices in health care.

It was recommended that the degree to which preferences over health states are informed should be studied in a larger study based on the methods implemented in this thesis. It was also argued that that reflection and deliberation may not help develop informed preferences and a potential method for obtaining better informed preferences is to provide participants with information on the consequences of ill health, such as activities, enjoyment, independence, relationships, dignity, and avoiding being a burden.

This thesis has extended the knowledge about the beliefs underlying health state values. It has been the second study to use a group-based reflection and deliberation exercise with a frequently used health descriptive system along with a frequently used choice-based valuation task. It was the first study to use a concurrent mixed methods design to explain the quantitative findings of the group deliberation. This thesis has developed a method to test whether preferences are informed. The findings, and limitations, from this thesis suggest that a larger study is warranted to investigate the extent that preferences are not informed and that methods which provide more and better information to respondents should be investigated. Pursuing these research directions can help secure the validity of methods used to value health.

References

Abreu, M.N., Siqueira, A.L. and Caiaffa, W.T. 2009. Ordinal logistic regression in epidemiological studies. *Revista de saude publica,* 43(1), 183-194.

Akunne, A.F., Bridges, J.F.P., Sanon, M. and Sauerborn, R. 2006. Comparison of individual and group valuation of health state scenarios across communities in West Africa. *Applied health economics and health policy,* 5(4), 261-268.

Al-Janabi, H., Flynn, T.N. and Coast, J. 2012. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Quality of Life Research,* 21(1), 167-176.

Ananda, J. and Herath, G. 2009. A critical review of multi-criteria decision making methods with special reference to forest management and planning. *Ecological Economics,* 68(10), 2535-2548.

Ananth, C.V. and Kleinbaum, D.G. 1997. Regression models for ordinal responses: a review of methods and applications. *Int J Epidemiol,* 26(6), 1323-1333.

Ariely, D., Loewenstein, G. and Prelec, D. 2003. "Coherent arbitrariness": stable demand curves without stable preferences. *Quarterly Journal of Economics,* 118, 73–105.

Arthur, S. and Nazroo, J. 2003. Designing fieldwork strategies and materials. *In* J. Ritchie and J. Lewis eds. *Qualitative research practice: a guide for social science students and researchers*. London, Sage. pp. 109-137.

Attema, A.E., Edelaar-Peeters, Y., Versteegh, M.M. and Stolk, E.A. 2013. Time trade-off: one methodology, different methods. *Eur J Health Econ,* 14 Suppl 1, S53-64.

Badia, X., Monserrat, S., Roset, M. and Herdman, M. 1999. Feasibility, validity and test-retest reliability of scaling methods for health states: the visual analogue scale and the time trade-off. *Quality of life research,* 8(4), 303-310.

Baker, R. and Robinson, A. 2004. Responses to standard gambles: are preferences 'well constructed'? *Health economics,* 13(1), 37-48.

Batterham, A.M. and Atkinson, G. 2005. How big does my sample need to be? A primer on the murky world of sample size estimation. *Phys Ther Sport,* 6(3), 153-163.

Beckerman, W. 2011. *Economics as applied ethics: value judgements in welfare economics*. Houndmills, Basingstoke, Hampshire ; New York: Palgrave Macmillan.

Belton, V. and Stewart, T.J. 2001. *Multiple Criteria Decision Analysis: An Integrated Approach*. Dordrecht: Kluwer Academic Publishers.

Bender, R. and Grouven, U. 1997. Ordinal logistic regression in medical research. *J Roy Coll Phys Lond,* 31(5), 546-551.

Bhattacharyya, A., Pattanaik, P.K. and Xu, Y.S. 2011. Choice, Internal Consistency and Rationality. *Economics and Philosophy,* 27(2), 123-149.

Booth, A. 2008. Unpacking your literature search toolbox: on search styles and tactics. *Health information and libraries journal,* 25(4), 313-317.

Brazier, J., Akehurst, R., Brennan, A., Dolan, P., Claxton, K., McCabe, C., Sculpher, M. and Tsuchyia, A. 2005. Should patients have a greater role in valuing health states? *Applied health economics and health policy,* 4(4), 201-208.

Brazier, J., Deverill, M. and Green, C. 1999a. A review of the use of health status measures in economic evaluation. *Journal of Health Services Research and Policy,* 4(3), 174-184.

Brazier, J., Deverill, M., Green, C., Harper, R. and Booth, A. 1999b. A review of the use of health status measures in economic evaluation. *Health Technology Assessment,* 3(9), iii-163.

Brazier, J. and McCabe, C. 2007. 'Is there a case for using visual analogue scale valuations in CUA' by Parkin and Devlin. A response: 'yes there is a case, but what does it add to ordinal data?'. *Health Econ,* 16(6), 645-647; discussion 649-651.

Brazier, J., Ratcliffe, J., Salomon, J.A. and Tsuchiya, A. 2007. *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford: Oxford University Press.

Broome, J. 2002. Measuring the Burden of Disease by Aggregating Well-being. *In* C. Murray, Salomon, J., Mathers, C. and Lopez, A. ed. *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications*. Geneva, World Health Organization. pp. 91–113.

Brouwer, W., Culyer, A.J., Van Exel, N.J.A. and Rutten, F.F.H. 2008. Welfarism vs. extra-welfarism. *Journal of health economics,* 27(2), 325-338.

Bryman, A. 2012. *Social research methods*. Fourth edition ed. Oxford: Oxford University Press.

Buckingham, K. and Devlin, N. 2006. A theoretical framework for TTO valuations of health. *Health Econ,* 15(10), 1149-1154.

Buitinga, L., Braakman-Jansen, L.M., Taal, E. and van de Laar, M.A. 2011. A computer Time Trade-Off: a feasible and reliable alternative for the interview Time Trade-Off in rheumatoid arthritis. *Clinical and experimental rheumatology,* 29(5), 783-789.

Burgess, J., Stirling, A., Clark, J., Davies, G., Eames, M., Staley, K. and Williamson, S. 2007. Deliberative mapping: a novel analytic-deliberative methodology to support contested science-policy decisions. *Public Understanding of Science,* 16(3), 299-322.

Cairns, J. 2006. Providing guidance to the NHS: The Scottish Medicines Consortium and the National Institute for Clinical Excellence compared. *Health Policy,* 76(2), 134-143.

Cicchetti, D.V. 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment,* 6(4), 284-290.

Coast, J. 1999. The appropriate uses of qualitative methods in health economics. *Health Econ,* 8(4), 345-353.

Coast, J., Flynn, T.N., Natarajan, L., Sproston, K., Lewis, J., Louviere, J.J. and Peters, T.J. 2008. Valuing the ICECAP capability index for older people. *Soc Sci Med,* 67(5), 874-882.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, N.J.: L. Erlbaum Associates.

Cohen, L., Manion, L. and Morrison, K. 2007. *Research methods in education*. 6th ed. New York: Routledge.

Connell, J., O'Cathain, A. and Brazier, J. 2014. Measuring quality of life in mental health: are we asking the right questions? *Soc Sci Med,* 120, 12-20.

Creswell, J.W. 2014. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Fourth ed. Thousands Oaks: Sage Publications.

Cribari-Neto, F. and Zeileis, A. 2010. Beta Regression in R. *J Stat Softw,* 34(2), 1-24.

Cubi-Molla, P., de Vries, J. and Devlin, N. 2014. A study of the relationship between health and subjective well-being in Parkinson's disease patients. *Value Health,* 17(4), 372-379.

Culyer, A.J. 1989. The normative economics of health care finance and provision. *Oxford Review of Economic Policy,* 5(1), 34-58.

Damschroder, L., Roberts, T., Goldstein, C., Miklosovic, M. and Ubel, P. 2005a. Trading people versus trading time: what is the difference? *Population health metrics,* 3(10).

Damschroder, L., Zikmund-Fisher, B.J. and Ubel, P.A. 2005b. The impact of considering adaptation in health state valuation. *Soc Sci Med,* 61(2), 267-277.

Damschroder, L., Zikmund-Fisher, B.J. and Ubel, P.A. 2008. Considering adaptation in preference elicitations. *Health Psychol,* 27(3), 394-399.

de Bekker-Grob, E.W., Ryan, M. and Gerard, K. 2012. Discrete choice experiments in health economics: a review of the literature. *Health Econ,* 21(2), 145-172.

Devlin, N., Hansen, P. and Selai, C. 2004. Understanding health state valuations: a qualitative analysis of respondents' comments. *Quality of life research,* 13(7), 1265-1277.

Devlin, N.J., Hansen, P., Kind, P. and Williams, A. 2003. Logical inconsistencies in survey respondents' health state valuations - a methodological challenge for estimating social tariffs. *Health Economics,* 12(7), 529-544.

Devlin, N.J. and Krabbe, P.F. 2013. The development of new research methods for the valuation of EQ-5D-5L. *Eur J Health Econ,* 14 Suppl 1, S1-3.

Devlin, N.J., Shah, K., Mulhern, B., Feng, Y., Tsuchiya, A. and Hout, B.v. 2014. Developing an EQ-5D-5L value set for England. *Health Economics Study Group, Winter 2014*. Sheffield.

Dimitrov, D.M. and Rumrill, P.D., Jr. 2003. Pretest-posttest designs and measurement of change. *Work,* 20(2), 159-165.

Dixon, S. 2011. Sheffield Dignity Questionnaire (SDQ). *PRO Newsletter,* 46, 9-11.

Dolan, P. 1996. The effect of experience of illness on health state valuations. *Journal of Clinical Epidemiology,* 49(5), 551-564.

Dolan, P. 1997a. Modeling Valuations for EuroQol Health States. *Medical care,* 35(11), 1095-1108.

Dolan, P. 1997b. The nature of individual preferences: a prologue to Johannesson, Jonsson and Karlsson. *Health Econ,* 6(1), 91-93.

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

Dolan, P., Gudex, C., Kind, P. and Williams, A. 1996b. Valuing health states: a comparison of methods. *J Health Econ,* 15(2), 209-231.

Dolan, P. and Kahneman, D. 2007. Interpretations Of Utility And Their Implications For The Valuation Of Health. *The Economic Journal,* 118(525), 215-234.

Dolan, P., Kavetsos, G. and Tsuchiya, A. 2013. Sick but satisfied: The impact of life and health satisfaction on choice between health scenarios. *Journal of Health Economics,* 32(4), 708-714.

Dolan, P. and Metcalfe, R. 2012. Valuing Health: A Brief Report on Subjective Well-Being versus Preferences. *Medical decision making,* 32(4), 578-582.

Dolan, P. and Olsen, J.A. 2002. *Distributing Health Care: Economic and ethical issues*. Oxford: Oxford University Press.

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

Dolan, P., Shaw, R., Tsuchiya, A. and Williams, A. 2005. QALY maximisation and people's preferences: A methodological review of the literature. *Health Economics,* 14(2), 197-208.

Drummond, M., Sculpher, M.J., Torrance, G., O'Brien, B. and Stoddart, G. 2005. *Methods for the Economic Evaluation of Health Care Programmes*. Third ed. Oxford: Oxford University Press.

Durlak, J.A. 2009. How to Select, Calculate, and Interpret Effect Sizes. *J Pediatr Psychol,* 34(9), 917-928.

Edwards, S., Lilford, R. and Kiauka, S. 1998. Different types of systematic review in health services research. *In* N. Black et al. eds. *Health Services Research Methods: a Guide to Best Practice*. London, BMJ Books. pp. 255-259.

Elsevier. 2015. *Scopus* [*online*]. Available at: http://www.scopus.com/.

Ericsson, K.A. and Simon, H.A. 1993. *Protocol Analysis - Rev'd Edition: Verbal Reports as Data*. Cambridge, MA: MIT Press.

Essink-Bot, M.-L., Stuifbergen, M.C., Meerding, W.-J., Looman, C.W.N. and Bonsel, G.J. 2007. Individual differences in the use of the response scale determine valuations of hypothetical health states: an empirical study. *BMC health services research,* 7(1), 62-62.

Fearon, J.D. 1998. Deliberation as Discussion. *In* J. Elster ed. *Deliberative Democracy*. Cambridge, Cambridge University Press. pp. 44-68.

Feeny, D., Furlong, W., Torrance, G.W., Goldsmith, C.H., Zhu, Z., DePauw, S., Denton, M. and Boyle, M. 2002. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Med Care,* 40(2), 113-128.

Ferrer-i-Carbonell, A. and Frijters, P. 2004. How important is methodology for the estimates of the determinants of happiness? *Econ J,* 114(497), 641-659.

Field, A. 2009. *Discovering Statistics Using SPSS*. Third ed: SAGE Publications Ltd.

Fischer, G.W. 1995. Range Sensitivity of Attribute Weight in Multiattribute Value Models. *Organizational Behavior and Human Decision Processes,* 62(3), 252-266.

Fischhoff, B. 1991. Value elicitation: is there anything in there? *Am Psychol,* 46(8), 835-847.

Flynn, T.N., Louviere, J.J., Peters, T.J. and Coast, J. 2010. Using discrete choice experiments to understand preferences for quality of life. Variance-scale heterogeneity matters. *Social Science & Medicine,* 70(12), 1957-1965.

Fox, J., Weisberg, S. and Fox, J. 2011. *An R companion to applied regression*. 2nd ed. Thousand Oaks, Calif.: SAGE Publications.

Gage, B.F., Cardinalli, A.B. and Owens, D.K. 1996. The effect of stroke and stroke prophylaxis with aspirin or warfarin on quality of life. *Archives of internal medicine,* 156(16), 1829-1836.

Gale, N.K., Heath, G., Cameron, E., Rashid, S. and Redwood, S. 2013. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *Bmc Med Res Methodol,* 13, 117.

Gandjour, A. 2010. Theoretical foundation of patient v. population preferences in calculating QALYs. *Medical decision making,* 30(4), E57-63.

Gold, M.R., Siegel, J.E., Russell, L.B. and Weinstein, M.C. 1996. *Cost-Effectiveness in Health and Medicine*. Oxford: Oxford University Press.

Goodin, R.E. and Niemeyer, S.J. 2003. When does deliberation begin? Internal reflection versus public discussion in deliberative democracy. *Political Studies,* 51(4), 627-649.

Google. 2015a. *Google* [*online*]. Available at: http://www.google.co.uk.

Google. 2015b. *Google Scholar* [*online*]. Available at: https://scholar.google.com/.

Grant, M.J. and Booth, A. 2009. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal,* 26(2), 91-108.

Green, C., Brazier, J. and Deverill, M. 2000. Valuing health-related quality of life. A review of health state valuation techniques. *Pharmacoeconomics,* 17(2), 151-165.

Gregory, R. and Slovic, P. 1997. A constructive approach to environmental valuation. *Ecological Economics,* 21(3), 175-181.

Griffin, J. 1986. *Well-being: its meaning, measurement, and moral importance*. Oxford Oxfordshire: Clarendon Press.

Gudex, C. 1994. Time trade-off user manual: props and self-completion methods. *Occasional Paper Series*. University of York.

Harrell, F.E. 2001. Regression Modeling Strategies With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York, NY, Springer.

Harsanyi, J.C. 1977. Morality and the Theory of Rational Behavior. *Social Research,* 44(4), 623-656.

Harsanyi, J.C. 1997. Utilities, preferences, and substantive goods. *Social Choice and Welfare,* 14(1), 129-145.

Hausman, D.M. 2006. Valuing health. *Philosophy public affairs,* 34(3), 246-274.

Hausman, D.M. 2010. Valuing health: a new proposal. *Health economics,* 19(3), 280-296.

Hausman, D.M. 2011. Mistakes about Preferences in the Social Sciences. *Philos Soc Sci,* 41(1), 3-25.

Hausman, D.M. 2012a. Measuring or Valuing Population Health: Some Conceptual Problems. *Public Health Ethics,*(i), 70-71.

Hausman, D.M. 2012b. *Preference, Value, Choice, and Welfare*. New York: Cambridge University Press.

Hausman, D.M. and McPherson, M.S. 2009. Preference Satisfaction and Welfare Economics. *Economics and Philosophy,* 25(1), 1-25.

Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., Bonsel, G. and Badia, X. 2011. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of life research,* 20(10), 1727-1736.

Hertzum, M., Hansen, K.D. and Andersen, H.H.K. 2009. Scrutinising usability evaluation: does thinking aloud affect behaviour and mental workload? *Behav Inform Technol,* 28(2), 165-181.

Hicks, S., Tinkler, L. and Allin, P. 2013. Measuring Subjective Well-Being and its Potential Role in Policy: Perspectives from the UK Office for National Statistics. *Social Indicators Research,* 114(1), 73-86.

Holyoak, K.J. and Simon, D. 1999. Bidirectional reasoning in decision making by constraint satisfaction. *J Exp Psychol Gen,* 128(1), 3-31.

Hougaard, J.L., Tjur, T. and Osterdal, L.P. 2012. On the meaningfulness of testing preference axioms in stated preference discrete choice experiments. *European Journal of Health Economics,* 13(4), 409-417.

Huang, I.B., Keisler, J. and Linkov, I. 2011. Multi-criteria decision analysis in environmental sciences: ten years of applications and trends. *Sci Total Environ,* 409(19), 3578-3594.

Janssen, M.F., Pickard, A.S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., Swinburn, P. and Busschbach, J. 2013. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Quality of life research,* 22(7), 1717-1727.

Kahneman, D. 1994. New Challenges to the Rationality Assumption. *Journal of Institutional and Theoretical Economics,* 150(1), 18-36.

Kitzinger, J. 1995. Qualitative Research: Introducing focus groups. *British Medical Journal,* 311(7000), 299-302.

Krabbe, P., Essink-Bot, M.L. and Bonsel, G.J. 1996. On the Equivalence of Collectively and Individually Collected Responses: Standard-gamble and Time-tradeoff Judgments of Health States. *Medical Decision Making,* 16(2), 120-132.

Krol, M., Sendi, P. and Brouwer, W. 2009. Breaking the Silence: Exploring the Potential Effects of Explicit Instructions on Incorporating Income and Leisure in TTO Exercises. *Value in Health,* 12(1), 172-180.

Lancsar, E. and Louviere, J. 2006. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? *Health economics,* 15(8), 797-811.

Lancsar, E. and Louviere, J. 2008. Conducting discrete choice experiments to inform healthcare decision making: A user's guide. *PharmacoEconomics,* 26(8), 661-677.

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

Lewis, J. 2003. Design Issues. *In* J. Ritchie and J. Lewis eds. *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. London, Sage. pp. 336-336.

Long, J.S. and Freese, J. 2006. *Regression models for categorical dependent variables using Stata.* 2nd ed. College Station, Texas: Stata Press.

Louviere, J.J. 1988. Conjoint-Analysis Modeling of Stated Preferences - a Review of Theory, Methods, Recent Developments and External Validity. *J Transp Econ Policy,* 22(1), 93-119.

Louviere, J.J., Flynn, T.N. and Carson, R.T. 2010. Discrete choice experiments are not conjoint analysis. *Journal of Choice Modelling,* 3(3), 57-72.

McFarlane, P.A., Pierratos, A., Bayoumi, A.M. and Redelmeier, D.A. 2007. Estimating preference scores in conventional and home nocturnal hemodialysis patients. *Clinical journal of the American Society of Nephrology : CJASN,* 2(3), 477-483.

McIntosh, C.N., Gorber, S.C., Bernier, J. and Berthelot, J.-m. 2007. Eliciting Canadian population preferences for health states using the Classifi cation and Measurement System of Functional Health (CLAMES). *Chronic Diseases in Canada,* 28(1), 29-41.

McTaggart-Cowan, H., Tsuchiya, A., O'Cathain, A. and Brazier, J. 2011. Understanding the effect of disease adaptation information on general population values for hypothetical health states. *Soc Sci Med,* 72(11), 1904-1912.

Merriam-Webster. 2015. *Merriam-Webster Online Dictionary* [*online*]. Available at: http://www.Merriam-Webster.com.

Miles, M.B. and Huberman, A.M. 1994. *Qualitative data analysis : an expanded sourcebook*. 2nd ed. Thousand Oaks: Sage Publications.

Monat, J.P. 2009. The benefits of global scaling in multi-criteria decision analysis. *Judgment and Decision Making,* 4(6), 492-508.

Morgan, D.L. 1996. Focus groups. *Annu Rev Sociol,* 22, 129-152.

Morris, S., Devlin, N.J. and Parkin, D. 2007. *Economic analysis in health care*. Chichester: John Wiley & Sons.

Morrison-Beedy, D., Cote-Arsenault, D. and Feinstein, N.F. 2001. Maximizing results with focus groups: moderator and analysis issues. *Appl Nurs Res,* 14(1), 48-53.

Mukuria, C. and Brazier, J. 2013. Valuing the EQ-5D and the SF-6D health states using subjective well-being: A secondary analysis of patient data. *Social Science & Medicine,* 77, 97-105.

Mukuria, C., Rowen, D., Peasgood, T. and Brazier, J. 2015. An empirical comparison of well-being measures used in the UK. *EEPRU Research Report 027*. Policy Research Unit in Economic Evaluation of Health and Care Interventions. Universities of Sheffield & York.

Mulhern, B., Longworth, L., Brazier, J., Rowen, D., Bansback, N., Devlin, N. and Tsuchiya, A. 2013. Binary choice health state valuation and mode of administration: head-to-head comparison of online and CAPI. *Value Health,* 16(1), 104-113.

Mulhern, B., Tsuchiya, A., Brazier, J. and Rowen, D. 2012. How do respondents perceive health state valuation exercises? A 'think aloud' study investigating Time Trade Off and Discrete Choice Experiments. *EuroQol Plenary 2012*.

Murray, C.J.L., Salomon, J.A., Mathers, C.D. and Lopez, A.D. 2002. Summary measures of population health: Conclusions and recommendations. *Summary measures of population health : Concepts, ethics, measurement and applications*. Geneva, Switzerland, World Health Organization. pp. 731-756.

Neumann, J.v. and Morgenstern, O. 2007. *Theory of Games and Economic Behavior*. Sixtieth-Anniversary Edition ed. Princeton: Princeton University Press.

NICE 2013a. Guide to the methods of technology appraisal. London.

NICE 2013b. Guide to the methods of technology appraisal. London.

NICE. 2015. *NICE consults on draft guidance on olaparib for ovarian, fallopian tube and peritoneal cancer* [*online*]. Available at: http://www.nice.org.uk/news/press-and-media/nice-consults-on-draft-guidance-on-olaparib-for-ovarian-fallopian-tube-and-peritoneal-cancer.

O'Cathain, A., Murphy, E. and Nicholl, J. 2007. Integration and Publications as Indicators of "Yield" From Mixed Methods Studies. *J Mix Method Res,* 1(2), 147-163.

O'Cathain, A., Murphy, E. and Nicholl, J. 2010. Three techniques for integrating data in mixed methods studies. *Bmj,* 341(November), 1147-1150.

OECD 2013. *OECD Guidelines on Measuring Subjective Well-being*. Paris: OECD Publishing.

Office for National Statistics 2011. Census: Aggregate data (England and Wales).

Office for National Statistics 2012. Analysis of experimental subjective well-being data from the Annual Population Survey, April to September 2011.

Oliver, A. 2000. Complete preferences over health states: a reply to the paper by Shiell et al. *Health Econ,* 9(8), 727-728.

Oliver, A. 2007. A Qualitative Analysis of the Lottery Equivalents Method. *Economics and Philosophy,* 23(02), 185-204.

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

Osch, S.v. and Stiggelbout, A.M. 2007. The construction of standard gamble utilities. *Health economics,* 40(April 2007), 31-40.

Ovid Technologies. 2015. *Ovid* [*online*]. Available at: http://ovidsp.ovid.com/.

Oxford University Press. 2015. *The Oxford English Dictionary* [*online*]. Available at: http://www.oed.com.

Palfreyman, S. and Mulhern, B. 2015. The psychometric performance of generic preference-based measures for patients with pressure ulcers. *Health and Quality of Life Outcomes,* 13.

Parkin, D. and Devlin, N. 2006. Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health economics,* 15(7), 653-664.

Patrick, D.L., Starks, H.E., Cain, K.C., Uhlmann, R.F. and Pearlman, R.A. 1994. Measuring Preferences for Health States Worse than Death. *Medical Decision Making,* 14(1), 9-18.

Payne, J.W., Bettman, J.R. and Johnson, E.J. 1992. Behavioral Decision Research - a Constructive Processing Perspective. *Annu Rev Psychol,* 43, 87-131.

Payne, J.W., Bettman, J.R. and Schkade, D.A. 1999. Measuring Constructed Preferences: Towards a Building Code. *Journal of Risk and Uncertainty*, 243-270.

Peacock, S., Richardson, J.R., Day, N. and Iezzi, A. 2004. The Assessment of Quality of Life (AQoL) II instrument. The effect of deliberation and alternative utility weights in a multi-attribute utility instrument. *Working Paper 143*. Monash University, Centre for Health Economics.

Peterson, B. and Harrell, F.E. 1990. Partial Proportional Odds Models for Ordinal Response Variables. *Journal of the Royal Statistical Society,* 39(2).

Pinheiro, J.C. and Bates, D.M. 2000. *Mixed-effects models in S and S-PLUS*. New York: Springer.

R Core Team 2015. R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing.

Rabin, R. and de Charro, F. 2001. EQ-5D: a measure of health status from the EuroQol Group. *Annals of medicine,* 33(5), 337-343.

Richardson, J., Iezzi, A., Khan, M.A. and Maxwell, A. 2014. Validity and Reliability of the Assessment of Quality of Life (AQoL)-8D Multi-Attribute Utility Instrument. *Patient-Patient Centered Outcomes Research,* 7(1), 85-96.

Richardson, J., Iezzi, A. and Maxwell, A. 2012. Cross-national comparison of twelve quality of life instruments: MIC Paper 1, Background, questions, instruments. Melbourne, Centre for Health Economics, Monash University.

Richardson, J., Khan, M.A., Iezzi, A. and Maxwell, A. 2015. Comparing and Explaining Differences in the Magnitude, Content, and Sensitivity of Utilities Predicted by the EQ-5D, SF-6D, HUI 3, 15D, QWB, and AQoL-8D Multiattribute Utility Instruments. *Med Decis Making,* 35(3), 276-291.

Ritchie, J. 2003. The applications of qualitative methods to social research. *In* J. Ritchie and J. Lewis eds. *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. London, Sage Publications. pp. 24-46.

Ritchie, J., Lewis, J. and Elam, G. 2003a. Designing and Selecting Samples. *In* J. Ritchie and J. Lewis eds. *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. London, Sage. pp. 77-108.

Ritchie, J. and Spencer, L. 1994. Qualitative data analysis for applied policy research. *In* Bryman and Burgess eds. *Analyzing qualitative data*. London, Routledge. pp. 173-194.

Ritchie, J., Spencer, L. and O'Connor, W. 2003b. Carrying out qualitative analysis. *In* J. Ritchie and J. Lewis eds. *Qualitative research practice: A guide for social science students and researchers*. London, SAGE publications. pp. 219-262.

Robinson, A., Dolan, P. and Williams, A. 1997. Valuing health status using VAS and TTO: what lies behind the numbers? *Social science & medicine,* 45(8), 1289-1297.

Robinson, A. and Parkin, D. 2002. Recognising diversity in public preferences: the use of preference sub-groups in cost-effectiveness analysis. A response to Sculpher and Gafni. *Health Econ,* 11(7), 649-651.

Robinson, S. 2011. Test-retest reliability of health state valuation techniques: the time trade off and person trade off. *Health Econ,* 20(11), 1379-1391.

Robinson, S. and Bryan, S. 2013. Does the Process of Deliberation Change Individuals' Health State Valuations? An Exploratory Study Using the Person Trade-Off Technique. *Value in Health,* 16(5), 806-813.

Ryan, M. and San Miguel, F. 2003. Revisiting the axiom of completeness in health care. *Health Economics,* 12(4), 295-307.

Ryan, M., Watson, V. and Entwistle, V. 2009. Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. *Health economics,* 336(July 2008), 321-336.

Sculpher, M. and Gafni, A. 2001. Recognizing diversity in public preferences: the use of preference sub-groups in cost-effectiveness analysis. *Health Econ,* 10(4), 317-324.

Sen, A. 1987. *Commodities and capabilities*. New Delhi: Oxford University Press.

Sen, A. 1993. Internal Consistency of Choice. *Econometrica,* 61(3), 495-521.

Sen, A. 2002. *Rationality and freedom*. Cambridge, MA: Harvard University Press.

Sendi, P. and Brouwer, W.B.F. 2005. Is silence golden? A test of the incorporation of the effects of ill-health on income and leisure in health state valuations. *Health Economics,* 14(6), 643-647.

Shaw, J.W., Johnson, J. and Coons, S.J. 2005. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Medical care,* 43(3), 203-220.

Sheffield 50+. 2014. *Sheffield 50+, The Voice of Older People in Sheffield* [*online*]. Available at: http://www.sheffield50plus.org.uk/.

Sheffield Community Information Service. 2015. *The Help Yourself Database* [*online*]. Available at: http://www.sheffieldhelpyourself.org.uk/.

Sheffield University of The Third Age. 2015. Available at: http://www.sheffieldu3a.org.uk/.

sheffieldu3a. 2015. *Links Newsletters* [*online*]. Available at: http://www.sheffieldu3a.org.uk/newsletter_home_page.html.

Sherman, S.N., Mrus, J.M., Yi, M.S., Feinberg, J. and Tsevat, J. 2006. How do patients with HIV/AIDS understand and respond to health value questions? *Journal of general internal medicine,* 21 Suppl 5(March), S56-61.

Shiell, A., Hawe, P. and Fletcher, M. 2003. Reliability of health utility measures and a test of values clarification. *Social science & medicine,* 56(7), 1531-1541.

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

Shrout, P.E. and Fleiss, J.L. 1979. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin,* 86(2), 420-428.

Simon, D., Krawczyk, D.C. and Holyoak, K.J. 2004. Construction of preferences by constraint satisfaction. *Psychological science,* 15(5), 331-336.

Skevington, S.M., Lotfy, M. and O'Connell, K.A. 2004. The World Health Organization's WHOQOL-BREF quality of life assessment: Psychometric properties and results of the international field trial - A report from the WHOQOL group. *Quality of Life Research,* 13(2), 299-310.

Slovic, P. 1995. The Construction of Preference. *Am Psychol,* 50(5), 364-371.

Smithson, M. and Verkuilen, J. 2006. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological methods,* 11(1), 54-71.

Snijders, T.A.B. and Bosker, R.J. 2012. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. 2nd ed. London: Sage.

Spencer, A. 2000. Testing the additive independence assumption in the QALY model. *Working paper*. London, Queen Mary, University of London.

Spencer, L., Ritchie, J. and O'Connor, W. 2003. Analysis: Practices, Principles, and Processes. *In* J. Ritchie and J. Lewis eds. *Qualitative research practice : a guide for social science students and researchers*. London, Sage Publications. pp. 199-218.

Stein, K., Ratcliffe, J., Round, A., Milne, R. and Brazier, J. 2006. Impact of discussion on preferences elicited in a group setting. *Health and Quality of Life Outcomes,* 4(22).

Sterne, J.A.C. and Smith, G.D. 2001. Sifting the evidence-what's wrong with significance tests? *British Medical Journal,* 322, 226-231.

Stevens, S.S. 1946. On the Theory of Scales of Measurement. *Science,* 103(2684), 677-680.

Stigler, G.J. and Becker, G.S. 1977. De Gustibus Non Est Disputandum. *Am Econ Rev,* 67(2), 76-90.

Stolk, E.A., Oppe, M., Scalone, L. and Krabbe, P.F. 2010. Discrete choice modeling for the quantification of health states: the case of the EQ-5D. *Value Health,* 13(8), 1005-1013.

Teixeira, P.A. and Schackman, B.R. 2008. Can urban methadone patients complete health utility assessments? *Patient Educ Couns,* 71(2), 302-307.

The WHOQOL group 1998. The World Health Organization Quality of Life assessment (WHOQOL): Development and general psychometric properties. *Social Science & Medicine,* 46(12), 1569-1585.

Thomson Reuters. 2015. *Web of Science* [*online*]. Available at: http://www.webofscience.com.

Tilling, C., Devlin, N., Tsuchiya, A. and Buckingham, K. 2010a. Protocols for time tradeoff valuations of health states worse than dead: a literature review. *Med Decis Making,* 30(5), 610-619.

Tilling, C., Krol, M., Tsuchiya, A., Brazier, J. and Brouwer, W. 2010b. In or Out? Income Losses in Health State Valuations: A Review. *Value in Health,* 13(2), 298-305.

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

Torrance, G. and Change 1986. Measurement of health state utilities for economic appraisal: a review. *Journal of Health Economics,* 5(1), 1-30.

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

Torrance, G.W. and Feeny, D. 1989. Utilities and quality-adjusted life years. *Int J Technol Assess Health Care,* 5(4), 559-575.

Tsevat, J. 2000. What do utilities measure? *Medical care,* 38(9), 60-164.

van der Pol, M. and Shiell, A. 2007. Extrinsic goals and time tradeoff. *Med Decis Making,* 27(4), 406-413.

van Hout, B., Janssen, M.F., Feng, Y.-S., Kohlmann, T., Busschbach, J., Golicki, D., Lloyd, A., Scalone, L., Kind, P. and Pickard, a.S. 2012. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value in health,* 15(5), 708-715.

van Nooten, F.E., Koolman, X. and Brouwer, W. 2009. The influence of subjective life expectancy on health state valuations using a 10 year TTO. *Health economics,* 18(5), 549-558.

van Osch, S.M. and Stiggelbout, A.M. 2005. Understanding VAS valuations: qualitative data on the cognitive process. *Quality of life research,* 14(10), 2171-2175.

Vaz, S., Falkmer, T., Passmore, A.E., Parsons, R. and Andreou, P. 2013. The Case for Using the Repeatability Coefficient When Calculating Test-Retest Reliability. *Plos One,* 8(9).

Venables, W.N. and Ripley, B.D. 2002. *Modern Applied Statistics with S*. Fourth ed. New York: Springer.

Versteegh, M.M., Attema, A.E., Oppe, M., Devlin, N.J. and Stolk, E.A. 2013. Time to tweak the TTO: results from a comparison of alternative specifications of the TTO. *Eur J Health Econ,* 14 Suppl 1, S43-51.

Vieira, E.S. and Gomes, J.A.N.F. 2009. A comparison of Scopus and Web of Science for a typical university. *Scientometrics,* 81(2), 587-600.

Walters, S.J. 2009. Quality of life outcomes in clinical trials and health-care evaluation a practical guide to analysis and interpretation. *Statistics in practice*. Chichester, John Wiley & Sons.

Walters, S.J. and Brazier, J.E. 2005. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Quality of life research,* 14(6), 1523-1532.

Ware, J.E. and Sherbourne, C.D. 1992. The MOS 36-Item Short-Form Health Survey (Sf-36). 1. Conceptual-Framework and Item Selection. *Medical Care,* 30(6), 473-483.

Wildman, D. 1995. Getting the most from paired-user testing. *interactions,* 2(3), 21-27.

Wittenberg, E., Halpern, E., Divi, N., Prosser, L.A., Araki, S.S. and Weeks, J.C. 2006. The effect of age, race and gender on preference scores for hypothetical health states. *Quality of life research,* 15(4), 645-653.

Wooldridge, J.M. 2009. *Introductory econometrics: a modern approach*. Fourth ed. Mason, OH: South-Western Cengage Learning.

Yee, T.W. and Wild, C.J. 1996. Vector generalized additive models. *Journal of the Royal Statistical Society Series B-Methodological,* 58(3), 481-493.

Appendix 1 Search strategy

Database used	**Keywords used**
Scopus Dates: 1960 to June 2015	((TITLE-ABS-KEY(preference)OR TITLE-ABS-KEY(preferences) OR TITLE-ABS-KEY(utility) OR TITLE-ABS-KEY(utilities)OR TITLE-ABS-KEY(values) OR TITLE-ABS-KEY(valuation))
	AND (TITLE-ABS-KEY(completeness) OR TITLE-ABS-KEY(reflection) OR TITLE-ABS-KEY(reflective) OR TITLE-ABS-KEY(deliberation) OR TITLE-ABS-KEY(deliberative) OR TITLE-ABS-KEY("group values") OR TITLE-ABS-KEY("group setting") OR TITLE-ABS-KEY("group valuation") OR TITLE-ABS-KEY("group responses") OR TITLE-ABS-KEY("collective judgment") OR ("adaptation exercise") **OR** TITLE-ABS-KEY("qualitative") OR TITLE-ABS-KEY("thought proces*") OR TITLE-ABS-KEY("think aloud") OR TITLE-ABS-KEY("think-aloud") OR TITLE-ABS-KEY("reasoning"))
	AND (TITLE-ABS-KEY("health state") OR TITLE-ABS-KEY("health states") OR TITLE-ABS-KEY("health-state") OR TITLE-ABS-KEY("health-states") OR TITLE-ABS-KEY("visual analogue scale") OR TITLE-ABS-KEY("Person trade-off") OR TITLE-ABS-KEY("time trade-off") OR TITLE-ABS-KEY("discrete choice experiment") OR TITLE-ABS-KEY("standard gamble")))
OVIDsp Dates: Econlit 1961 to June 2015 Medline 1946 to June 2015 Psyinfo 1987 to June 2015	("health state" or "health states" or "health-state" or "health-states" or "visual analogue scale"or "person trade-off" or "time trade-off" or "discrete choice experiment" or "standard gamble")
	AND (completeness OR reflection OR reflective OR deliberation OR deliberative OR "group values" OR "group setting" OR "group valuation" OR "group responses" OR "collective judgment" OR "adaptation exercise" **OR** "qualitative" OR "thought proces*" OR "think aloud" OR "think-aloud" OR "reasoning")
	AND (Preference or Preferences or Utility or Utilities or Values or Valuation)

Appendix 2 Documents relating to ethics approval of study 1

Information Sheet

The University Of Sheffield.

1. **Research Project Title**

Investigating how people evaluate health states

2. **Invitation**

You are being invited to take part in a research project. Before you decide whether to participate, it is important to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. ***Our contact information is at the end of this information sheet.***

3. **What is the project's purpose?**

The NHS needs to know the value that people place on different medical conditions in order to make decisions about what services to provide. The aim of this project is to find out what people think about when considering the value of health to them.

4. **Why have I been chosen?**

We wish to interview about twenty members of the general public for this project. You have been chosen as a member of the general public. Unfortunately, we cannot include people who cannot read or speak English or are under the age of 18.

5. **Do I have to take part?**

It is up to you to decide whether or not to take part. If you do decide to take part, you will be given a consent form to sign. You can still withdraw at anytime, you do not have to give a reason and there will be no penalty. No more information will be collected if you decide to withdraw from the study.

6. **What do I have to do?**

You will be invited at your convenience to come to Regent Court near West Street at the University of Sheffield. You will be given a consent form to sign to say that you are happy to undertake the interview.

We will conduct an interview that should last about one hour. You will be asked to complete a questionnaire and to say what you were thinking about as you were completing the questionnaire. The questionnaire asks for your opinion on different health outcomes to see which one you value the most. This will involve imagining living with certain health conditions for different lengths of time before you die. You will also be asked to fill in a short questionnaire about your current health.

7. **Will I be recorded, and how will the recorded media be used?**

We will make audio recordings and take notes during the interview. This study has to be audio recorded and unfortunately we cannot accept participants who do not want to be audio recorded. The audio recordings made during the interview will be transcribed and all

personal references removed. No one outside the project team will be allowed access to the original recordings. The original recordings will be destroyed after the project is over (in approximately two years' time).

All recordings and transcripts will be stored in a safe and secure location on the campus of the University of Sheffield, where only members of the research team have access.

9. **What are the possible benefits of taking part?**

This project will help the NHS better understand which aspects of health matter most for people. This may allow policy makers to better understand which health care treatments are most important to people.

10. **What are the possible disadvantages and risks of taking part?**

Thinking about bad states of health can be difficult and upsetting. In considering the questions we ask, you may have to imagine living with some severe health conditions. You may also have to imagine making difficult choices about the value of life with different health conditions and to consider whether life with certain conditions is worth living.

This study also requires you to travel to the University of Sheffield. You will be compensated with a **£**10 gift voucher for your time and travel expenses.

11. **What if something goes wrong?**

In first instance you may contact the PhD student Milad Karimi (contact details are at the end). You may also contact the lead supervisor, John Brazier, at:

The School of Health and Related Research (ScHARR)
The University of Sheffield
Regent Court, 30 Regent Street
Sheffield, S1 4DA

12. **Will my taking part in this project be kept confidential?**

The information that we collect will be kept strictly confidential. You will not be identified or identifiable in any reports or publications. The audio recordings will be transcribed and all references to personal details removed. The researchers will use the anonymised transcripts, not the actual voice recordings. After the project is finished, the audio recordings will be destroyed. The anonymised transcripts will be archived on the university campus.

13. **What will happen to the results of the research project?**

The results of this project will be used for a subsequent research project and form part of a PhD thesis. Some of the data may be used for journal articles or conference presentations. All results that are made public will be anonymised, by removing any words that could identify you, and you will not be identifiable in any publications.

14. **Who is organising and funding the research?**

The research is part of a PhD study that is funded by a faculty studentship from Faculty of Medicine, Dentistry and Health of the University of Sheffield.

15. **Who has ethically reviewed the project?**

The School of Health and Health Related Research of the University of Sheffield has reviewed and approved this project.

16. **Contact for further information**

You may contact Milad Karimi:

Telephone: 0114 222 6382

Email: mkarimi@sheffield.ac.uk

Post:
Innovation Centre,
217 Portobello,
S1 4DP,
Sheffield, South Yorkshire

Thank you for considering this research project

Date: 11-12-2012

Ethics approval study 1 (continued) – Consent form

UNIVERSITY OF SHEFFIELD

TITLE OF RESEARCH PROJECT: **INVESTIGATING HOW PEOPLE EVALUATE HEALTH STATES**

NAME OF RESEARCHER: MILAD KARIMI

PARTICIPANT IDENTIFICATION NUMBER:

PLEASE INITIAL BOX

1. I CONFIRM THAT I HAVE READ AND UNDERSTOOD THE INFORMATION SHEET DATED ***[11-12-2012]*** EXPLAINING THE ABOVE RESEARCH PROJECT AND I HAVE HAD THE OPPORTUNITY TO ASK QUESTIONS ABOUT THE PROJECT. ☐
2. I UNDERSTAND THAT MY PARTICIPATION IS ENTIRELY VOLUNTARY AND THAT I AM FREE TO WITHDRAW AT ANY TIME WITHOUT GIVING ANY REASON AND WITHOUT THERE BEING ANY PENALTY. ☐
3. I CONSENT TO BEING AUDIO RECORDED. ☐
4. 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. ☐
5. I AGREE FOR THE ANONYMISED DATA COLLECTED FROM ME TO BE USED IN FUTURE RESEARCH. ☐
6. I AGREE TO TAKE PART IN THE ABOVE RESEARCH PROJECT. ☐

_____	_____	_____
NAME OF PARTICIPANT	DATE	SIGNATURE
_____	_____	_____
LEAD RESEARCHER	DATE	SIGNATURE

CONTACT INFORMATION:

MILAD KARIMI
0114 222 6382
MKARIMI@SHEFFIELD.AC.UK

INNOVATION CENTRE
217 PORTOBELLO, SHEFFIELD
SOUTH YORKSHIRE, S1 4DP

Ethics approval study 1 (continued) – Approval letter

The University Of Sheffield.

Kirsty Woodhead
Ethics Committee Administrator

Regent Court
30 Regent Street
Sheffield S1 4DA
Telephone: +44 (0) 114 2225453
Fax: +44 (0) 114 272 4095 (non confidential)
Email: k.woodhead@sheffield.ac.uk

Our ref: 0606/KW

14 December 2012

Milad Karimi
ScHARR

Dear Milad

Investigating how people evaluate health states.

Thank you for submitting the above research project for approval by the ScHARR Research Ethics Committee. On behalf of the University Chair of Ethics who reviewed your project, I am pleased to inform you that on 14 December 2012 the project was approved on ethics grounds, on the basis that you will adhere to the documents that you submitted for ethics review.

The research must be conducted within the requirements of the hosting/employing organisation or the organisation where the research is being undertaken. You are also required to ensure that you meet any research ethics and governance requirements in the country in which you are researching. It is your responsibility to find out what these are.

If during the course of the project you need to deviate significantly from the documents you submitted for review, please inform me since written approval will be required. Please also inform me should you decide to terminate the project prematurely.

Yours sincerely

Kirsty Woodhead
Ethics Committee Administrator

Appendix 3 Discrete choice task warm-up

Car purchase choice

If you were offered the option between car A and B below which one would you choose?

Car A	Car B
£ 7,000 purchase price 5 star safety £500 per year maintenance cost Large boot (trunk)	£ 5,000 purchase price 4 star safety rating £600 per year maintenance cost Small boot (trunk)
☐ Choose car A	☐ Choose car B

Appendix 4 Background questionnaire

1. Age:

☐ 18-25
☐ 25-34
☐ 35-44
☐ 45-54
☐ 55+

2. Gender

☐ Male
☐ Female

3. What is your main professional activity?

☐ Full-time student
☐ Employed or self-employed
☐ Retired
☐ Unemployed

4. Marital status:

☐ Married or living with partner
☐ Single or never married
☐ Divorced, separated or widowed

5. Do you or someone close to you have any previous experience of serious illness? (You may select multiple boxes)

☐ No
☐ Yourself
☐ Family
☐ Friends
☐ Other

6. Do you have children?

☐ Yes
☐ No

7. What is the highest education level that you have completed?

☐ Other qualification
☐ O-level/GCSE
☐ A-level
☐ Diploma
☐ First Degree
☐ University Higher Degree

8. Do you consider yourself religious?

☐ Yes
☐ No

Appendix 5 EQ-5D-5L questionnaire

Under each heading, please tick the ONE box that best describes your health TODAY

MOBILITY

I have no problems in walking about ❑
I have slight problems in walking about ❑
I have moderate problems in walking about ❑
I have severe problems in walking about ❑
I am unable to walk about ❑

SELF-CARE

I have no problems washing or dressing myself ❑
I have slight problems washing or dressing myself ❑
I have moderate problems washing or dressing myself ❑
I have severe problems washing or dressing myself ❑
I am unable to wash or dress myself ❑

USUAL ACTIVITIES *(e.g. work, study, housework, family or leisure activities)*

I have no problems doing my usual activities ❑
I have slight problems doing my usual activities ❑
I have moderate problems doing my usual activities ❑
I have severe problems doing my usual activities ❑
I am unable to do my usual activities ❑

PAIN / DISCOMFORT

I have no pain or discomfort ❑
I have slight pain or discomfort ❑
I have moderate pain or discomfort ❑
I have severe pain or discomfort ❑
I have extreme pain or discomfort ❑

ANXIETY / DEPRESSION

I am not anxious or depressed ❑
I am slightly anxious or depressed ❑
I am moderately anxious or depressed ❑
I am severely anxious or depressed ❑
I am extremely anxious or depressed ❑

Appendix 6 EQ-5D-5L health states used in study one

Task	DCE State A	DCE State B
1	10 years in 33243 • moderate problems in walking about • moderate problems washing or dressing yourself • slight problems doing usual activities • severe pain or discomfort • moderately anxious or depressed	10 years in 33234 • moderate problems in walking about • moderate problems washing or dressing yourself • slight problems doing usual activities • moderate pain or discomfort • severely anxious or depressed
2	10 years in 21221 • slight problems in walking about • no problems washing or dressing yourself • slight problems doing usual activities • slight pain or discomfort • not anxious or depressed	10 years in 12131 • no problems in walking about • slight problems washing or dressing yourself • no problems doing usual activities • moderate pain or discomfort • not anxious or depressed
3	10 years in 13323 • no problems in walking about • moderate problems washing or dressing yourself • moderate problems doing usual activities • no pain or discomfort • moderately anxious or depressed	10 years in 31332 • moderate problems in walking about • slight problems washing or dressing yourself • slight problems doing usual activities • slight pain or discomfort • slightly anxious or depressed
4	8 years in 43312 • severe problems in walking about • moderate problems washing or dressing yourself • moderate problems doing usual activities • no pain or discomfort • moderately anxious or depressed	10 years in 33411 • moderate problems in walking about • moderate problems washing or dressing yourself • severe problems doing usual activities • slight pain or discomfort • slightly anxious or depressed
5	5 years in 34454 • moderate problems in walking about • severe problems washing or dressing yourself • severe problems doing your usual activities • extreme pain or discomfort • severely anxious or depressed	5 years in 43544 • moderate problems in walking about • severe problems washing or dressing yourself • unable to do your usual activities • severe pain or discomfort • severely anxious or depressed

Task	DCE State A	DCE State B
6	10 years in 23211 • slight problems in walking about • moderate problems washing or dressing yourself • slight problems doing your usual activities • no pain or discomfort • not anxious or depressed	10 years in 12311 • no problems in walking about • slight problems washing or dressing yourself • slight problems doing your usual activities • no pain or discomfort • slightly anxious or depressed
7	5 years in 33341 • severe problems in walking about • moderate problems washing or dressing yourself • moderate problems doing your usual activities • severe pain or discomfort • severely anxious or depressed	5 years in 53321 • unable to walk about • moderate problems washing or dressing yourself • moderate problems doing your usual activities • slight pain or discomfort • slightly anxious or depressed
8	8 years in 22432 • slight problems in walking about • slight problems washing or dressing yourself • severe problems doing usual activities • moderate pain or discomfort • slightly anxious or depressed	10 years in 22233 • slight problems in walking about • slight problems washing or dressing yourself • slight problems doing usual activities • moderate pain or discomfort • moderately anxious or depressed

TTO task	TTO state	
1	13321	• no problems in walking about • moderate problems washing or dressing yourself • moderate problems doing your usual activities • slight pain or discomfort • not anxious or depressed
2	13443	• no problems in walking about • moderate problems washing or dressing yourself • severe problems doing your usual activities • severe pain or discomfort • moderately anxious or depressed
3	54435	• unable to walk about • severe problems washing or dressing yourself • severe problems doing your usual activities • moderate pain or discomfort • extremely anxious or depressed
4	31212	• moderate problems walking about • no problems washing or dressing yourself • slight problems doing your usual activities • no pain or discomfort • slightly anxious or depressed

Appendix 7 Discrete choice experiment sample format

DCE 1

If you had to choose between the scenarios below, which would you choose?

Life A	Life B
You live 10 years with:	You live 10 years with:
• You have moderate problems in walking about • You have moderate problems washing or dressing yourself • You have slight problems doing your usual activities • You have severe pain or discomfort • You are moderately anxious or depressed	• You have moderate problems in walking about • You have moderate problems washing or dressing yourself • You have slight problems doing your usual activities • You have moderate pain or discomfort • You are severely anxious or depressed
Then you die	Then you die
☐ Choose life A	☐ Choose life B

Appendix 8 TTO board

Appendix 9 Documents relating to ethics approval study 2 – Information sheet

UNIVERSITY OF SHEFFIELD

Information Sheet

The University Of Sheffield.

1. **Research Project Title**

Investigating the effect of reflection and deliberation on health state values

2. **Invitation**

You are being invited to take part in a research project. Before you decide whether to participate, it is important to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. ***Our contact information is at the end of this information sheet.***

3. **What is the project's purpose?**

The decisions made by the NHS about what services to provide are partly informed by people's views of the relative desirability of different health conditions. The aim of this project is to find out how people determine how (un)desirable different health conditions are, and to compare different methods for determining which aspects of health are most important to people. Your participation will help researchers understand how people decide which aspects of health and which treatments are most important to them.

4. **Why have I been chosen?**

We wish to recruit about 70 members of the general public for this project. You have been chosen as a member of the general public. Unfortunately, we cannot include people who cannot read or speak English or are under the age of 18.

5. **Do I have to take part?**

It is up to you to decide whether or not to take part. If you do decide to take part, you will be asked to sign a consent form. You can still withdraw at anytime, you do not have to give a reason and there will be no penalty. No more information will be collected if you decide to withdraw from the study.

6. **What do I have to do?**

You will be invited to come to Regent Court near West Street at the University of Sheffield for a group meeting. This meeting will be with 5 to 7 other participants. You will be given a consent form to sign to say that you are happy to undertake the group meeting.

You will take part in a meeting that should last about two hours in total. During this meeting you will complete several tasks, some individually and some as a group. You will be asked to value several health outcomes individually. This will involve imagining living with certain health conditions for different lengths of time before you die. You will then be asked to discuss your thoughts about the task with other group members. You will also be asked to fill in a short questionnaire about your current health and a short background form.

7. **Will I be recorded, and how will the recorded media be used?**

We will make audio recordings and take notes during the group meeting. This study has to be audio recorded and unfortunately we cannot accept participants who do not want to be

audio recorded. The audio recordings made during the interview will be transcribed and all personal references removed. No one outside the project team will be allowed access to the original recordings. The original recordings will be destroyed after the project is over (in approximately one years' time).

All recordings and transcripts will be stored in a safe and secure location on the campus of the University of Sheffield, where only members of the research team have access.

8. **What are the possible benefits of taking part?**

This project will help the NHS better understand which aspects of health matter most for people. This may allow policy makers to better understand which health care treatments are most important to people.

9. **What are the possible disadvantages and risks of taking part?**

Thinking about serious states of ill health can be difficult and upsetting. In considering the questions we ask, you may have to imagine living with some severe health conditions. You may also have to imagine making difficult choices about the value of life with different health conditions and to consider whether life with certain conditions is worth living. You will also be asked to share your thoughts and opinions with the other members in the group, not all of whom will necessarily agree with your views. It can be difficult to discuss upsetting topics such as ill health in a group, particularly if there are disagreements within the group. However, it is important to remember that there is no right or wrong answer, and that your own view and perspective is valued in this study. You are also always free to take a break or to withdraw from the study at any time.

10. **What if something goes wrong?**

We do not anticipate that anything will go wrong. Should you be unhappy with any part of your involvement in the study, please contact the lead supervisor, John Brazier, at:

Tel: (+44) (0)114 222 0726
E-mail: j.e.brazier@sheffield.ac.uk
The School of Health and Related Research (ScHARR)
The University of Sheffield
Regent Court, 30 Regent Street
Sheffield, S1 4DA

11. **Will my taking part in this project be kept confidential?**

The information that we collect will be kept strictly confidential. You will not be identified or identifiable in any reports or publications. The audio recordings will be transcribed and all references to personal details removed. The researchers will use the anonymised transcripts, not the actual voice recordings. After the project is finished, the audio recordings will be destroyed. The anonymised transcripts will be archived on the university campus for two years after the completion of the study.

12. **What will happen to the results of the research project?**

The results of this project will form part of a PhD thesis. Some of the data may be used for journal articles or conference presentations. All results that are made public will be anonymised, by removing any words that could identify you, and you will not be identifiable in any publications.

13. **Who is organising and funding the research?**

The research is part of a PhD study that is funded by a faculty studentship from Faculty of Medicine, Dentistry and Health of the University of Sheffield.

14. **Who has ethically reviewed the project?**

The ethics committee of the School of Health and Related Research of the University of Sheffield has reviewed and approved this project.

15. **Contact for further information**

You may contact Milad Karimi:

Telephone: 0114 222 6382

Email: mkarimi@sheffield.ac.uk

Address:
Innovation Centre,
217 Portobello,
S1 4DP,
Sheffield, South Yorkshire

Thank you for considering this research project

Ethics approval study 2 (continued) – Consent form

UNIVERSITY OF SHEFFIELD

The University Of Sheffield.

TITLE OF RESEARCH PROJECT: **INVESTIGATING THE EFFECT OF REFLECTION AND DELIBERATION ON HEALTH STATE VALUES**
NAME OF RESEARCHER: MILAD KARIMI

PARTICIPANT IDENTIFICATION NUMBER:

PLEASE INITIAL EACH BOX

1. I CONFIRM THAT I HAVE READ AND UNDERSTOOD THE INFORMATION SHEET, VERSION NUMBER **V2**, EXPLAINING THE ABOVE RESEARCH PROJECT AND I HAVE HAD THE OPPORTUNITY TO ASK QUESTIONS ABOUT THE PROJECT. ☐
2. I UNDERSTAND THAT MY PARTICIPATION IS ENTIRELY VOLUNTARY AND THAT I AM FREE TO WITHDRAW AT ANY TIME WITHOUT GIVING ANY REASON AND WITHOUT THERE BEING ANY PENALTY. ☐
3. I CONSENT TO BEING AUDIO RECORDED. ☐
4. 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. ☐
5. I AGREE FOR THE ANONYMISED DATA COLLECTED FROM ME TO BE USED IN FUTURE RESEARCH. ☐
6. I AGREE TO KEEP THE INFORMATION DISCUSSED BY OTHER PARTICIPANTS CONFIDENTIAL AND TO NOT DISCLOSE ANYTHING WITH PEOPLE OUTSIDE OF THE GROUP. ☐
7. I AGREE TO TAKE PART IN THE ABOVE RESEARCH PROJECT. ☐

_____	_____	_____
NAME OF PARTICIPANT	DATE	SIGNATURE
_____	_____	_____
LEAD RESEARCHER	DATE	SIGNATURE

CONTACT INFORMATION:
MILAD KARIMI
0114 222 6382
MKARIMI@SHEFFIELD.AC.UK
INNOVATION CENTRE
217 PORTOBELLO, SHEFFIELD
SOUTH YORKSHIRE, S1 4DP

Ethics approval study 2 (continued) – Approval letter

The University Of Sheffield.

Kirsty Woodhead
Ethics Committee Administrator

Regent Court
30 Regent Street
Sheffield S1 4DA
Telephone: +44 (0) 114 2225453
Fax: +44 (0) 114 272 4095 (non confidential)
Email: k.woodhead@sheffield.ac.uk

Our ref: 0687/KW

23 October 2013

Milad Karimi
ScHARR

Dear Milad

Investigating the effect of reflection and deliberation on health state values

Thank you for submitting the above research project for approval by the ScHARR Research Ethics Committee. On behalf of the University Chair of Ethics who reviewed your project, I am pleased to inform you that on 23 October 2013 the project was approved on ethics grounds, on the basis that you will adhere to the documents that you submitted for ethics review.

The research must be conducted within the requirements of the hosting/employing organisation or the organisation where the research is being undertaken. You are also required to ensure that you meet any research ethics and governance requirements in the country in which you are researching. It is your responsibility to find out what these are.

If during the course of the project you need to deviate significantly from the documents you submitted for review, please inform me since written approval will be required. Please also inform me should you decide to terminate the project prematurely.

Yours sincerely

Kirsty Woodhead
Ethics Committee Administrator

Appendix 10 Topic guide for facilitator study two

Stage	**Prompts**
Welcome • Information sheet and consent form • Turn on recorder	1. Welcome 2. Aim and study reasons 3. Four stages of study 4. Confidentiality 5. Group discussion rules (no consensus) 6. Sign consent forms 7. Everyone introduce themselves and how they travelled to the group meeting
TTO • Hand out TTO booklet and ranking envelope • Collect TTO booklet	1. Stage is focused on valuing health 2. Fill in EQ-5D-5L for own health 3. Rank 8 health states in envelope (20 years in state as it is) 4. Practice TTO description 5. Describe WTD possibility 6. Ask to complete six real health states valuation
MCDA scoring • Hand out MCDA booklet	1. Explain MCDA – focus on life with health state 2. Explain six criteria 3. Explain scoring method 4. Ask to score health states on six criteria (wait for last page)
MCDA weights • Leave MCDA booklet on table	1. Explain criteria weights 2. Explain how to weigh (most important is 100)
Break • Hand out cold refreshments	1. Break for 10 minutes, refreshments, bathroom directions
Group discussion	1. Aim (reasoning, knowledge, experience) 2. Rules (no consensus, allowed to leave, confidentiality) 3. Criteria 1 discussion (which health state highest, lowest, anyone agree/disagree, what was important in scoring) 4. Other criteria in turn 5. If time, compare MCDA and TTO briefly
TTO • Hand out 2nd TTO booklet • Hand out vouchers	1. Reminder of TTO instructions 2. Ask to fill in all health states and background form 3. Sign finance forms
Debrief	1. Answer any questions

Appendix 11 First TTO booklet sample pages

The University Of Sheffield.

School of Health and Related Research

Form 1 – Valuing health and illness

Participant ID:_____

Date: _____

Thank you for taking part in this study!

Section [A] – Your health state

Under each heading, please tick the **ONE** box that best describes your health **TODAY**

1. MOBILITY

I have no problems in walking about .....q
I have slight problems in walking about .....q
I have moderate problems in walking about .....q
I have severe problems in walking about .....q
I am unable to walk about .....q

2. SELF-CARE

I have no problems washing or dressing myself .....q
I have slight problems washing or dressing myself .....q
I have moderate problems washing or dressing myself .....q
I have severe problems washing or dressing myself .....q
I am unable to wash or dress myself .....q

3. USUAL ACTIVITIES *(e.g. work, study, housework, family or leisure activities)*

I have no problems doing my usual activities .....q
I have slight problems doing my usual activities .....q
I have moderate problems doing my usual activities .....q
I have severe problems doing my usual activities .....q
I am unable to do my usual activities .....q

4. PAIN / DISCOMFORT

I have no pain or discomfort .....q
I have slight pain or discomfort .....q
I have moderate pain or discomfort .....q
I have severe pain or discomfort .....q
I have extreme pain or discomfort .....q

5. ANXIETY / DEPRESSION

I am not anxious or depressed .....q
I am slightly anxious or depressed .....q
I am moderately anxious or depressed .....q
I am severely anxious or depressed .....q
I am extremely anxious or depressed .....q

Section [B] Valuation exercise

[B1] Ranking exercise

The envelope contains eight health states. Please rank them in order of severity, starting with the best at the top. You may rank them equally, just cross out the number for the equal ranked health states. Please write down your ranking of the health states below.

Rank	Health state label
1.	
2.	
3.	
4.	
5.	
6.	
7.	
8.	

Section [B] Valuation exercise

[B2] Valuation instructions

You are going to be asked to make a choice between two lives in different health states and for different lengths of time. The health states will be the states that you have just ranked.

For Choice A we would like you to imagine that you will live in the health state for the number of years that it mentions, and then you will die. For Choice B we would like you to imagine that you will live in the health state for 20 years, and then you will die.

We would like you to consider how many years in Choice A would be equivalent to 20 years in Choice B. Choice B will change with every question. From now on, imagine that you yourself are in these states, and that they would last for the number of years mentioned without change.

Section [B] Valuation exercise

[B3] Practice question 1A

If you prefer 0 years of choice A over 20 years in choice B, please answer question practice 1B over the page.
Please put an A against all cases where you are CONFIDENT that you would choose choice A.
Please put a B against all cases where you are CONFIDENT that you would choose choice B.
Please put an = against all cases where you cannot choose between choice A and choice B.
Choice A and B are described on the next page.

Choice A	Would you choose A, B or are they the same?	Choice B
20 Years		20 Years
19 Years		20 Years
18 Years		20 Years
17 Years		20 Years
16 Years		20 Years
15 Years		20 Years
14 Years		20 Years
13 Years		20 Years
12 Years		20 Years
11 Years		20 Years
10 Years		20 Years
9 Years		20 Years
8 Years		20 Years
7 Years		20 Years
6 Years		20 Years
5 Years		20 Years
4 Years		20 Years
3 Years		20 Years
2 Years		20 Years
1 Year		20 Years
0 Years		20 Years

If you have filled in A in the last row, please go to page 8. Otherwise, please go to page 10.

Section [B] Valuation exercise

[B3] Practice question 1A

Choice A

- I have no problems in walking about
- I have no problems washing or dressing myself
- I have no problems doing my usual activities
- I have no pain or discomfort
- I am not anxious or depressed

Choice B

- I have slight problems in walking about
- I have slight problems washing or dressing myself
- I have slight problems doing my usual activities
- I have slight pain or discomfort
- I am slightly anxious or depressed

Section [B] Valuation exercise

Practice question 1B

Please put an A against all cases where you are CONFIDENT that you would choose choice A.
Please put a B against all cases where you are CONFIDENT that you would choose choice B.
Please put an = against all cases where you cannot choose between choice A and choice B.
Choice A and B are described on the next page.

Choice A		Would you choose A, B or are they the same?	Choice B
State 1	State 2		
20 Years	0 Years		Dead
19 Years	1 Year		Dead
18 Years	2 Years		Dead
17 Years	3 Years		Dead
16 Years	4 Years		Dead
15 Years	5 Years		Dead
14 Years	6 Years		Dead
13 Years	7 Years		Dead
12 Years	8 Years		Dead
11 Years	9 Years		Dead
10 Years	10 Years		Dead
9 Years	11 Years		Dead
8 Years	12 Years		Dead
7 Years	13 Years		Dead
6 Years	14 Years		Dead
5 Years	15 Years		Dead
4 Years	16 Years		Dead
3 Years	17 Years		Dead
2 Years	18 Years		Dead
1 Year	19 Years		Dead

Please go to page 10.

Section [B] Valuation exercise

Practice question 1B

Choice A

State 1

- I have slight problems in walking about
- I have slight problems washing or dressing myself
- I have slight problems doing my usual activities
- I have slight pain or discomfort
- I am slightly anxious or depressed

State 2

- I have no problems in walking about
- I have no problems washing or dressing myself
- I have no problems doing my usual activities
- I have no pain or discomfort
- I am not anxious or depressed

Choice B

Dead

Appendix 12 Second TTO booklet – sample pages

The University Of Sheffield.

School of Health and Related Research

Form 3 – Valuing health and illness

Participant ID:_____

Date: _____

Thank you for taking part in this study!

Section [A] Valuation exercise

[A1] Valuation instructions

You are going to be asked to make a choice between two lives in different health states and for different lengths of time.

For Choice A we would like you to imagine that you will live in the health state for the number of years that it mentions, and then you will die. For Choice B we would like you to imagine that you will live in the health state for 20 years, and then you will die.

We would like you to consider how many years in Choice A would be equivalent to 20 years in Choice B. Choice B will change with every question. From now on, imagine that you yourself are in these states, and that they would last for the number of years mentioned without change.

Section [A] Valuation exercise

[A1] Question 1A

If you prefer 0 years of choice A over 20 years in choice B, please answer question 1B over the page.

Please put an A against all cases where you are CONFIDENT that you would choose choice A.
Please put a B against all cases where you are CONFIDENT that you would choose choice B.
Please put an = against all cases where you cannot choose between choice A and choice B.
Choice A and B are described on the next page.

Choice A	Would you choose A, B or are they the same?	Choice B
20 Years		20 Years
19 Years		20 Years
18 Years		20 Years
17 Years		20 Years
16 Years		20 Years
15 Years		20 Years
14 Years		20 Years
13 Years		20 Years
12 Years		20 Years
11 Years		20 Years
10 Years		20 Years
9 Years		20 Years
8 Years		20 Years
7 Years		20 Years
6 Years		20 Years
5 Years		20 Years
4 Years		20 Years
3 Years		20 Years
2 Years		20 Years
1 Year		20 Years
0 Years		20 Years

If you have filled in A in the last row, please go to page 6. Otherwise, please go to page 8.

Section [A] Valuation exercise

Question 1A

Choice A

- I have no problems in walking about
- I have no problems washing or dressing myself
- I have no problems doing my usual activities
- I have no pain or discomfort
- I am not anxious or depressed

Choice B

- I have moderate problems in walking about
- I have slight problems washing or dressing myself
- I have moderate problems doing my usual activities
- I have slight pain or discomfort
- I am slightly anxious or depressed

5

Section [A] Valuation exercise

Question 1B

Please put an A against all cases where you are CONFIDENT that you would choose choice A.
Please put a B against all cases where you are CONFIDENT that you would choose choice B.
Please put an = against all cases where you cannot choose between choice A and choice B.
Choice A and B are described on the next page.

Choice A		Would you choose A, B or are they the same?	Choice B
State 1	State 2		
20 Years	0 Years		Dead
19 Years	1 Year		Dead
18 Years	2 Years		Dead
17 Years	3 Years		Dead
16 Years	4 Years		Dead
15 Years	5 Years		Dead
14 Years	6 Years		Dead
13 Years	7 Years		Dead
12 Years	8 Years		Dead
11 Years	9 Years		Dead
10 Years	10 Years		Dead
9 Years	11 Years		Dead
8 Years	12 Years		Dead
7 Years	13 Years		Dead
6 Years	14 Years		Dead
5 Years	15 Years		Dead
4 Years	16 Years		Dead
3 Years	17 Years		Dead
2 Years	18 Years		Dead
1 Year	19 Years		Dead

Please go to page 8.

Section [A] Valuation exercise

Question 1B

Choice A

State 1

- I have moderate problems in walking about
- I have slight problems washing or dressing myself
- I have moderate problems doing my usual activities
- I have slight pain or discomfort
- I am slightly anxious or depressed

State 2

- I have no problems in walking about
- I have no problems washing or dressing myself
- I have no problems doing my usual activities
- I have no pain or discomfort
- I am not anxious or depressed

Choice B

Dead

Section [A] Valuation exercise

Background questionnaire

1. What is your age: _____

2. Gender:

Male ..... ☐
Female..... ☐

3. What is your main professional activity?

Full-time student..... ☐
Employed or self-employed ..... ☐
Retired..... ☐
Unemployed..... ☐
Housework ..... ☐
Long-term sick..... ☐

4. Marital status:

Married or living with partner..... ☐
Single or never married..... ☐
Divorced, separated or widowed ..... ☐

5. Have you experienced serious illness?

in you yourself……Yes ☐……..No ☐

in your family……Yes ☐……..No ☐

in caring for others……Yes ☐……..No ☐

6. Do you have children?

Yes..... ☐
No..... ☐

7. Please select each education level that you have completed:

Other qualification..... ☐
O-level/GCSE..... ☐
A-level..... ☐
Diploma ..... ☐
First Degree ..... ☐
University Higher Degree..... ☐

8. Do you consider yourself religious?

Yes ..... ☐
No ..... ☐

Please hand in your booklet.
Thank you for taking part in this study!

Appendix 13 MCDA booklet

School of Health and Related Research

Form 2 – Considering life with an illness

Participant ID:_____

Date: _____

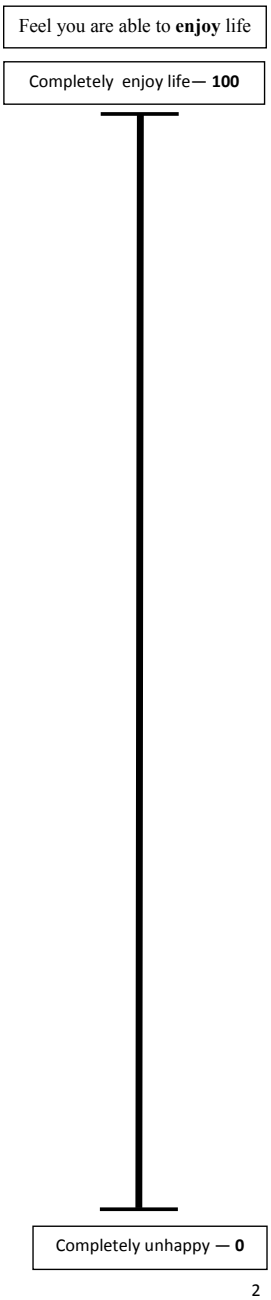
Thank you for taking part in this study!

Criteria 1 - Would you feel you are able to **enjoy** life?

We want you to rate the following health states with respect to the criteria stated above.

Imagine that each state would last 20 years and then you would die.

Please put a score from 0, meaning completely unhappy, to 100, meaning completely able to enjoy life, for each health state in the second column.



Label	Health states	Score
GY	I have **severe** problems in walking about I have **severe** problems washing or dressing myself I am **unable to** do my usual activities I have **moderate** pain or discomfort I am **extremely** anxious or depressed	
ED	I have **no** problems in walking about I have **no** problems washing or dressing myself I have **no** problems doing my usual activities I have **no** pain or discomfort I am **not** anxious or depressed	
YM	I have **no** problems in walking about I have **no** problems washing or dressing myself I have **moderate** problems doing my usual activities I have **moderate** pain or discomfort I am **not** anxious or depressed	
IR	I have **moderate** problems in walking about I have **slight** problems washing or dressing myself I have **moderate** problems doing my usual activities I have **slight** pain or discomfort I am **slightly** anxious or depressed	
YC	I have **moderate** problems in walking about I have **no** problems washing or dressing myself I have **no** problems doing my usual activities I have **moderate** pain or discomfort I am **not** anxious or depressed	
AU	**Being dead**	
NA	I have **severe** problems in walking about I have **severe** problems washing or dressing myself I am **unable to do** my usual activities I have **extreme** pain or discomfort I am **moderately** anxious or depressed	
UI	I have **no** problems in walking about I have **no** problems washing or dressing myself I have **moderate** problems doing my usual activities I have **moderate** pain or discomfort I am **severely** anxious or depressed	

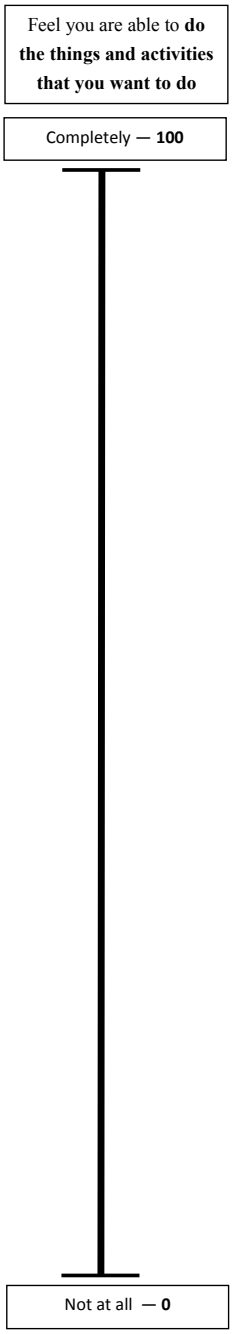
Criteria 2 - Would you feel able to **do the things and activities that you want to do**?

We want you to rate the following health states with respect to the criteria stated above.

Imagine that each state would last 20 years and then you would die.

Please put a score from 0, meaning not at all able to do the things and activities that you want to do, to 100, meaning completely able to do the things and activities that you want to do, for each health state in the second column.

Label	Health states	Score
GY	I have **severe** problems in walking about I have **severe** problems washing or dressing myself I am **unable to** do my usual activities I have **moderate** pain or discomfort I am **extremely** anxious or depressed	
ED	I have **no** problems in walking about I have **no** problems washing or dressing myself I have **no** problems doing my usual activities I have **no** pain or discomfort I am **not** anxious or depressed	
YM	I have **no** problems in walking about I have **no** problems washing or dressing myself I have **moderate** problems doing my usual activities I have **moderate** pain or discomfort I am **not** anxious or depressed	
IR	I have **moderate** problems in walking about I have **slight** problems washing or dressing myself I have **moderate** problems doing my usual activities I have **slight** pain or discomfort I am **slightly** anxious or depressed	
YC	I have **moderate** problems in walking about I have **no** problems washing or dressing myself I have **no** problems doing my usual activities I have **moderate** pain or discomfort I am **not** anxious or depressed	
AU	**Being dead**	
NA	I have **severe** problems in walking about I have **severe** problems washing or dressing myself I am **unable to do** my usual activities I have **extreme** pain or discomfort I am **moderately** anxious or depressed	
UI	I have **no** problems in walking about I have **no** problems washing or dressing myself I have **moderate** problems doing my usual activities I have **moderate** pain or discomfort I am **severely** anxious or depressed	



3

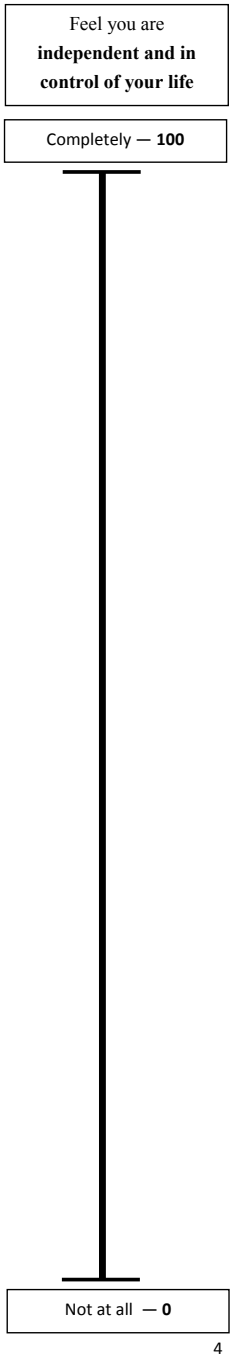
Criteria 3 - Would you feel **independent and in control of your life?**

We want you to rate the following health states with respect to the criteria stated above.

Imagine that each state would last 20 years and then you would die.

Please put a score from 0, meaning not at all independent and in control of your life, to 100, meaning completely independent and in control of your life, for each health state in the second column.

Label	Health states	Score
GY	I have **severe** problems in walking about I have **severe** problems washing or dressing myself I am **unable to** do my usual activities I have **moderate** pain or discomfort I am **extremely** anxious or depressed	
ED	I have **no** problems in walking about I have **no** problems washing or dressing myself I have **no** problems doing my usual activities I have **no** pain or discomfort I am **not** anxious or depressed	
YM	I have **no** problems in walking about I have **no** problems washing or dressing myself I have **moderate** problems doing my usual activities I have **moderate** pain or discomfort I am **not** anxious or depressed	
IR	I have **moderate** problems in walking about I have **slight** problems washing or dressing myself I have **moderate** problems doing my usual activities I have **slight** pain or discomfort I am **slightly** anxious or depressed	
YC	I have **moderate** problems in walking about I have **no** problems washing or dressing myself I have **no** problems doing my usual activities I have **moderate** pain or discomfort I am **not** anxious or depressed	
AU	**Being dead**	
NA	I have **severe** problems in walking about I have **severe** problems washing or dressing myself I am **unable to do** my usual activities I have **extreme** pain or discomfort I am **moderately** anxious or depressed	
UI	I have **no** problems in walking about I have **no** problems washing or dressing myself I have **moderate** problems doing my usual activities I have **moderate** pain or discomfort I am **severely** anxious or depressed	

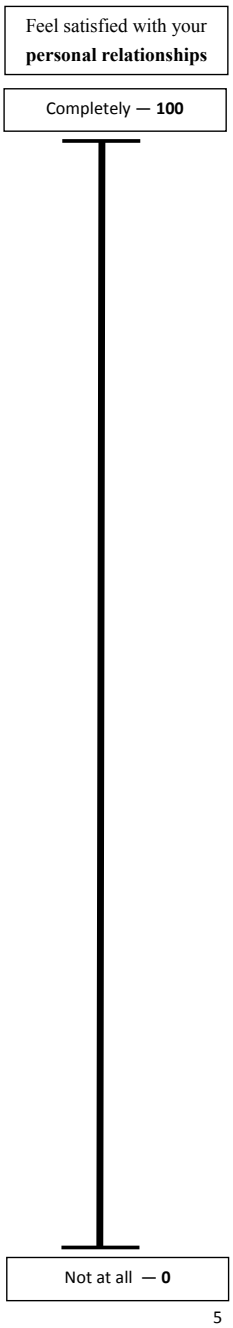


Criteria 4 - Would you feel satisfied with your **personal relationships?**

We want you to rate the following health states with respect to the criteria stated above.

Imagine that each state would last 20 years and then you would die.

Please put a score from 0, meaning not at all happy with your relationships, to 100, meaning completely happy with your relationships, for each health state in the second column.



Label	Health states	Score
GY	I have **severe** problems in walking about I have **severe** problems washing or dressing myself I am **unable to** do my usual activities I have **moderate** pain or discomfort I am **extremely** anxious or depressed	
ED	I have **no** problems in walking about I have **no** problems washing or dressing myself I have **no** problems doing my usual activities I have **no** pain or discomfort I am **not** anxious or depressed	
YM	I have **no** problems in walking about I have **no** problems washing or dressing myself I have **moderate** problems doing my usual activities I have **moderate** pain or discomfort I am **not** anxious or depressed	
IR	I have **moderate** problems in walking about I have **slight** problems washing or dressing myself I have **moderate** problems doing my usual activities I have **slight** pain or discomfort I am **slightly** anxious or depressed	
YC	I have **moderate** problems in walking about I have **no** problems washing or dressing myself I have **no** problems doing my usual activities I have **moderate** pain or discomfort I am **not** anxious or depressed	
AU	**Being dead**	
NA	I have **severe** problems in walking about I have **severe** problems washing or dressing myself I am **unable to do** my usual activities I have **extreme** pain or discomfort I am **moderately** anxious or depressed	
UI	I have **no** problems in walking about I have **no** problems washing or dressing myself I have **moderate** problems doing my usual activities I have **moderate** pain or discomfort I am **severely** anxious or depressed	

5

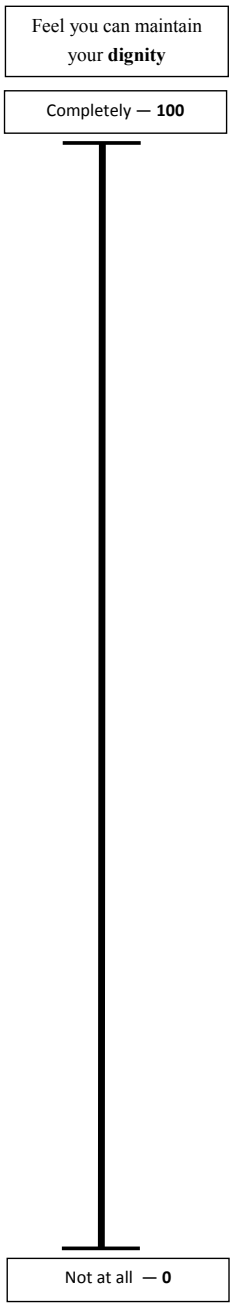
Criteria 5 - Would you feel you can maintain your **dignity?**

We want you to rate the following health states with respect to the criteria stated above.

Imagine that each state would last 20 years and then you would die.

Please put a score from 0, meaning not at all able to maintain your dignity, to 100, meaning completely able to maintain your dignity, for each health state in the second column.

Label	Health states	Score
GY	I have **severe** problems in walking about I have **severe** problems washing or dressing myself I am **unable to** do my usual activities I have **moderate** pain or discomfort I am **extremely** anxious or depressed	
ED	I have **no** problems in walking about I have **no** problems washing or dressing myself I have **no** problems doing my usual activities I have **no** pain or discomfort I am **not** anxious or depressed	
YM	I have **no** problems in walking about I have **no** problems washing or dressing myself I have **moderate** problems doing my usual activities I have **moderate** pain or discomfort I am **not** anxious or depressed	
IR	I have **moderate** problems in walking about I have **slight** problems washing or dressing myself I have **moderate** problems doing my usual activities I have **slight** pain or discomfort I am **slightly** anxious or depressed	
YC	I have **moderate** problems in walking about I have **no** problems washing or dressing myself I have **no** problems doing my usual activities I have **moderate** pain or discomfort I am **not** anxious or depressed	
AU	**Being dead**	
NA	I have **severe** problems in walking about I have **severe** problems washing or dressing myself I am **unable to do** my usual activities I have **extreme** pain or discomfort I am **moderately** anxious or depressed	
UI	I have **no** problems in walking about I have **no** problems washing or dressing myself I have **moderate** problems doing my usual activities I have **moderate** pain or discomfort I am **severely** anxious or depressed	



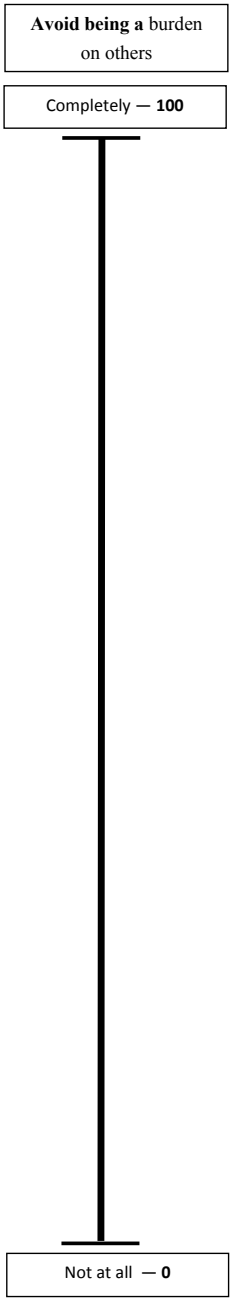
Criteria 6 - Would you be able to **avoid being a burden** on others?

We want you to rate the following health states with respect to the criteria stated above.

Imagine that each state would last 20 years and then you would die.

Please put a score from 0, meaning not at all able to avoid being a burden, to 100, meaning completely able to avoid being a burden, for each health state in the second column.

Label	Health states	Score
GY	I have **severe** problems in walking about I have **severe** problems washing or dressing myself I am **unable to** do my usual activities I have **moderate** pain or discomfort I am **extremely** anxious or depressed	
ED	I have **no** problems in walking about I have **no** problems washing or dressing myself I have **no** problems doing my usual activities I have **no** pain or discomfort I am **not** anxious or depressed	
YM	I have **no** problems in walking about I have **no** problems washing or dressing myself I have **moderate** problems doing my usual activities I have **moderate** pain or discomfort I am **not** anxious or depressed	
IR	I have **moderate** problems in walking about I have **slight** problems washing or dressing myself I have **moderate** problems doing my usual activities I have **slight** pain or discomfort I am **slightly** anxious or depressed	
YC	I have **moderate** problems in walking about I have **no** problems washing or dressing myself I have **no** problems doing my usual activities I have **moderate** pain or discomfort I am **not** anxious or depressed	
AU	**Being dead**	
NA	I have **severe** problems in walking about I have **severe** problems washing or dressing myself I am **unable to do** my usual activities I have **extreme** pain or discomfort I am **moderately** anxious or depressed	
UI	I have **no** problems in walking about I have **no** problems washing or dressing myself I have **moderate** problems doing my usual activities I have **moderate** pain or discomfort I am **severely** anxious or depressed	



7

Criteria weighting

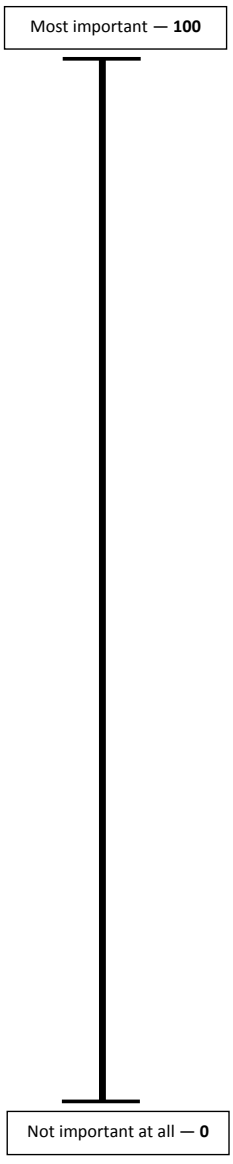
We want to know how important the criteria are for you.

From the list please mark on the second column what you think is the most important criteria. Now imagine that the most important criteria is worth a 100 points. Please put a number on the other criteria to reflect their relative importance. For example, if another criteria is about half as important, please write 50.

If you feel a criteria is not at all important or relevant, you can write 0.

All numbers should be between 0 and 100.

Criteria	Check if most important criteria	Value (out of 100)
Being able to enjoy life		
Being able to do the things and activities that you want to do		
Being independent and in control of your life		
Being satisfied with your relationships		
Maintaining your dignity		
Avoiding being a burden on others		



8

a

Appendix 14 Replication of analysis of Chapter 6 with US transformation

Health state EQ-5D-5L profile	Before Value (SD)	After value (SD)	Change (SD)	P-value change	Cohen's d	ICC
44535	0.204 (0.31)	0.165 (0.42)	-0.039 (0.11)	0.52	0.11	0.63
31131	0.736 (0.22)	0.754 (0.2)	0.018 (-0.01)	0.63	0.09	0.77
32322	0.659 (0.25)	0.634 (0.29)	-0.024 (0.04)	0.56	0.09	0.72
11331	0.75 (0.24)	0.761 (0.2)	0.011 (-0.04)	0.52	0.05	0.62
11334	0.376 (0.39)	0.335 (0.45)	-0.041 (0.05)	0.28	0.10	0.78
44553	0.136 (0.43)	0.168 (0.42)	0.032 (-0.01)	0.40	0.08	0.68

Appendix 15 Linear mixed model anova results for six health states

Model	**Fixed effects**	**Random effects**	**Change in deviance (p-value)**					
			Health state 44535	Health state 31131	Health state 32322	Health state 11331	Health state 11334	Health state 44553
1	-	Group and individual	-	-	-	-	-	-
2	Time	Group and individual	1 (0.29)	1 (0.34)	1 (0.44)	0 (0.66)	0 (0.53)	0 (0.51)
3	Time	Group, individual, and Time	5 (0.33)	1 (0.94)	1 (0.85)	7 (0.14)	2 (0.68)	3 (0.61)
4	Time Gender Main activity Marital Children Education Religious Own EQ-5D value Illness experience	Group and individual	4 (0.89)	10 (0.26)	12 (0.14)	5 (0.77)	12 (0.14)	12 (0.17)

Appendix 16 Linear mixed model estimates of model two for six health states

Random effect	**Variance**					
	Health state 44535	Health state 31131	Health state 32322	Health state 11331	Health state 11334	Health state 44553
Groups	0.01	0.01	0.00	0.00	0.04	0.03
Individual	0.16	0.03	0.08	0.03	0.18	0.19
Residual	0.09	0.01	0.03	0.02	0.06	0.10
Fixed effects	**Estimate (standard error)**					
	Health state 44535	Health state 31131	Health state 32322	Health state 11331	Health state 11334	Health state 44553
Intercept	0.07 (0.07)	0.74 (0.03)	0.64 (0.04)	0.75 (0.03)	0.31 (0.09)	-0.02 (0.09)
Valuation after group meeting	-0.06 (0.05)	0.02 (0.02)	-0.02 (0.03)	0.01 (0.03)	-0.03 (0.05)	0.04 (0.06)

Appendix 17 Similarity of themes found across groups in study two

Group	Count of disagreements	Count of comments about personal nature of valuation	Count of comments about opinion changing	Count of comments about adaptation not being considered	Count of comments indicating surprise	Count of comments indicating uncertainty	Count of comments indicating grouping of health states
1	3	4	1	-	-	2	-
2	4	2	4	2	-	1	5
3	3	4	2	-	-	2	4
4	3	3	-	-	-	3	5
5	4	4	-	-	-	4	6
6	3	5	-	-	-	5	7
7	2	1	1	-	-	-	13
8	1	1	-	-	-	5	11
9	1	4	2	-	-	6	12
10	-	1	-	-	-	5	7
11	3	2	-	1	3	1	8
12	5	2	-	-	-	6	12
13	1	6	1	-	-	2	-

Appendix 18 Ethics for study three

The University Of Sheffield.

SCHOOL OF HEALTH AND RELATED RESEARCH.

ScHARR

ELLIE NICOLSON
ETHICS ADMINISTRATOR

REGENT COURT
30 REGENT STREET
SHEFFIELD S1 4DA

TELEPHONE: +44 (0) 114 222 25446
FAX: +44 (0) 114 272 4095 (NON CONFIDENTIAL)
EMAIL: E.L.NICOLSON@SHEFFIELD.AC.UK

16 SEPTEMBER 2015

ScHARR

DEAR MILAD,

PROJECT TITLE: ARE HEALTH STATE PREFERENCES RATIONAL?

I AM PLEASED TO INFORM YOU THAT YOUR PROJECT HAS BEEN CLASSED AS 'LOW RISK' SO YOU CAN PROCEED WITH YOUR RESEARCH. THE RESEARCH MUST BE CONDUCTED WITHIN THE REQUIREMENTS OF THE HOSTING/EMPLOYING ORGANISATION OR THE ORGANISATION WHERE THE RESEARCH IS BEING UNDERTAKEN.

I HAVE RECEIVED A COPY OF YOUR DECLARATION TOGETHER WITH YOUR CONFIRMATION FOR RESEARCH THAT DOES NOT INVOLVE HUMAN PARTICIPANTS AND THAT YOU WILL BE UNDERTAKING RESEARCH WHICH INVOLVES ANALYSIS OF ALREADY EXISTING DATA ('SECONDARY DATA').

YOURS SINCERELY

E. Nicolson

ELLIE NICOLSON
ON BEHALF OF THE ScHARR RESEARCH ETHICS COMMITTEE

Appendix 19 Probability calculations for Relationships

Health state	**Probability for each outcome level**				**Expected value**	**Ranking**
	4 (worst)	3	2	1 (best)		
31131	0.001	0.049	0.352	0.598	1.453	1
11331	0.001	0.059	0.370	0.569	1.493	2
32322	0.016	0.230	0.442	0.312	1.949	3
44553	0.059	0.420	0.367	0.153	2.385	4
11334	0.094	0.494	0.309	0.104	2.577	5
44545	0.200	0.576	0.184	0.039	2.938	6

Appendix 20 Probability calculations for Independence

Health state	**Probability for each outcome level**				**Expected value**	**Ranking**
	4 (worst)	3	2	1 (best)		
31131	0.002	0.025	0.285	0.689	1.339	1
11331	0.006	0.075	0.463	0.456	1.632	2
11334	0.029	0.235	0.582	0.154	2.140	3
32322	0.036	0.271	0.573	0.121	2.222	4
44553	0.202	0.626	0.171	0.002	3.027	5
44545	0.259	0.638	0.102	0.001	3.156	6

Appendix 21 Probability calculations for Activities

Health state	Probability for each outcome level					Expected value	Ranking
	5 (worst)	4	3	2	1 (best)		
31131	0.018	0.046	0.137	0.199	0.599	1.685	1
11331	0.088	0.163	0.310	0.242	0.197	2.703	2
32322	0.175	0.257	0.348	0.169	0.051	3.336	3
11334	0.239	0.302	0.326	0.115	0.018	3.628	4
44553	0.672	0.270	0.056	0.002	0.000	4.613	5
44545	0.720	0.241	0.038	0.001	0.000	4.681	6

Appendix 22 Expected estimate for Enjoyment

Health state	Expected value	Ranking
31131	0.774	1
11331	0.739	2
32322	0.593	3
44553	0.382	4
11334	0.297	5
44545	0.167	6

Appendix 23 Probability calculations for Burden

Health state	**Probability for each outcome level**					**Expected value**	**Ranking**
	5 (worst)	4	3	2	1 (best)		
31131	0.001	0.006	0.039	0.245	0.709	1.345	1
11331	0.003	0.022	0.093	0.358	0.523	1.624	2
32322	0.032	0.130	0.264	0.411	0.163	2.458	3
11334	0.094	0.273	0.328	0.265	0.041	3.114	4
44553	0.226	0.420	0.259	0.091	0.004	3.772	5
44545	0.385	0.454	0.140	0.021	0.000	4.204	6

Appendix 24 Probability calculations for Dignity

EQ-5D-5L state	Possible 3L states[1]	Probability of 3L state[1]	Probability for each outcome level			Expected value for 3L state	Expected value for 5L state	Ranking from Best
			3 (worst)	2	1 (best)			
11331	11221	1.00	0.00	0.03	0.97	1.03	1.03	1
31131	21121	1.00	0.01	0.12	0.87	1.14	1.14	2
11334	11222	0.51	0.00	0.08	0.91	1.07	1.155	3
	11223	0.49	0.02	0.19	0.80	1.24		
32322	21211	0.01	0.01	0.13	0.87	1.16	1.40	4
	21212	0.03	0.04	0.30	0.66	1.38		
	21221	0.03	0.02	0.18	0.80	1.22		
	21222	0.11	0.07	0.39	0.54	1.53		
	22211	0.03	0.02	0.20	0.79	1.25		
	22212	0.13	0.08	0.41	0.51	1.57		
	22221	0.14	0.03	0.27	0.69	1.32		
	22222	0.53	0.13	0.48	0.39	1.74		
44553	22332	0.70	0.31	0.55	0.15	2.18	2.24	5
	23332	0.23	0.36	0.53	0.11	2.25		
	32332	0.05	0.34	0.54	0.12	2.22		
	33332	0.02	0.40	0.51	0.09	2.31		
44535	22323	0.70	0.39	0.52	0.09	2.3	2.38	6
	23323	0.23	0.45	0.48	0.06	2.37		
	32323	0.05	0.43	0.50	0.07	2.36		
	33323	0.02	0.49	0.46	0.05	2.44		

1: (van Hout et al., 2012)