RM03/JH11

The relative societal value of health gains to different beneficiaries

Final report

31 January 2008

Paul Dolan, Imperial College London
Richard Edlin, University of Leeds
Aki Tsuchiya, University of Sheffield

On behalf of the full research team:

Chris Armitage, University of Sheffield
Amy Hukin, University of Sheffield
John Brazier, University of Sheffield
Rachel Ibbotson, Sheffield Hallam University
Stirling Bryan, University of Birmingham
Dick Eiser, University of Sheffield
Jan Abel Olsen, University of Tromsø

List of abbreviations

This study considers both individual and societal valuations of health and to aid comparison it uses societal analogues to the standard individual terms, as below.

Individual health valuations		**Societal health valuations**	
QOL	Quality of life	FHA	Value of 100% health to an adult
		SHA	Value of 25% health to an adult
		FHC	Value of 100% health to a child
		SHC	Value of 25% health to a child
QALY	Quality adjusted life year	AHYE	Adult healthy year equivalent

Other abbreviations are as below:

BHPS	British Household Panel Survey
CEA	Cost-effectiveness analysis
CES	Constant elasticity of substitution
HE/FE	Higher education/further education
IO	Interview only (Phase 2 variant)
PAT	Persuasive argumentation theory
MRS	Marginal rate of substitution
MRSA	Methicillin-Resistance *Streptococcus Aureus*
RI	Resource intensive (Phase 2 variant)
SCT	Social comparison theories
SWF	Social welfare function

EXECUTIVE SUMMARY

Background

Since health is a function of both length of life and quality of life, the quality-adjusted life-year (QALY) has been developed in an attempt to combine the value of these two into a single index number. QALYs are increasingly being used in the evaluation of health care interventions and are used by the National Institute for Health and Clinical Excellence (NICE) for health technology coverage decisions in England and Wales.

The de facto standard in economic evaluations is that “a QALY is a QALY is a QALY” i.e. that a given health benefit is valued equally regardless of who gets it. However, there is growing evidence (albeit often from small convenience samples) that this assumption is not a good representation of societal preferences. It does not, however, allow us to estimate ‘equity weights’ for QALYs i.e. weights based on the relative societal value of health gains to different beneficiaries.

Aims

Against this background, the main aim of this project was to generate a set of equity weights for QALYs from the preferences of the UK general population. Our specific goal was to elicit preferences from members of the general public that can be fed into a social welfare function (SWF).

The SWF allows for some QALYs to be weighted more highly than others for two reasons. The first is the degree to which inequalities in health matter: we may be willing to sacrifice some overall health for a more equal distribution of health (essentially by giving a greater weight to lower levels of health). Inequalities are defined in terms of lifetime health. The second is the degree to which responsibility for health matters: we may be willing to give greater weight to lower levels of health that are the result of ‘bad luck’ as compared to ‘bad choices’.

Before we could do this, we need, first, to determine the relevant attributes and levels over which to elicit preferences and, second, to develop methods which facilitate the elicitation of stable preferences.

Determining the relevant attributes and levels

Relevant attributes are ones that are potentially relevant to policy making and ones that the general public think should apply to macro level resource allocation decisions of the kind faced by NICE. Relevant levels of a relevant attribute are ones over which the general public are willing to make a clear distinction e.g. age may matter when the choice is between a child and an adult but not when it is a choice between a 40 year-old and a 50 year-old.

Based on the results of focus groups with 57 members of the general public and a survey of 172 NHS employees, and consistent with our general SWF approach, the following attributes were the focus of investigation: the timing of ill health, the severity of ill health and responsibility.

There was no real consensus about when the timing of ill health mattered for policy except in the case of children versus adults, which are the focus of our comparisons.

For quality of life, values of 25% and 50% were chosen for comparison with full health (100%) as these points on the health scale are clearly distinguishable from one another.

Responsibility is categorised in three ways: NHS responsibility and no patient responsibility; no NHS responsibility and some patient responsibility; and no NHS responsibility and no patient responsibility. These categories capture concerns for safety and recognise that few conditions are entirely due to individual responsibility.

Additional questions replicated the responsibility questions with specific illness labels. No NHS responsibility and some patient responsibility was relabelled as ‘obesity-related conditions’; NHS responsibility and no patient responsibility was relabelled as ‘MRSA-related complications’; and no NHS and no patient responsibility was relabelled as either ‘genetic conditions’ or ‘workplace hazards’.

We also added condition rarity as an additional attribute at the request of NICE, who were interested in preferences relevant to ultra-orphan conditions. The question asked respondents to compare an extremely rare condition with a rare one.

Methods for eliciting stable preferences

We are more likely to elicit a stable set of preferences if we first elicit general beliefs and attitudes. In addition to this, we considered the degree to which more stable preferences might be elicited through a discussion group prior to individual interviews.

Based on a comparison of 56 interviews conducted after a group discussion and 232 interviews conducted ‘cold’, we decided to proceed with the latter variant. The discussion groups did not appear to have much effect on preferences at the interview stage and they were more costly. The prior elicitation of general attitudes serves as a more efficient ‘warm-up’.

Methods for generating equity weights

Respondents make multiple pairwise choices, from a citizen perspective, between hypothetical scenarios involving two groups of patients with differing health and/or non-health characteristics. Respondents are asked to "select the one that you would recommend NICE to choose". Ordinal responses are then converted into cardinal equity weights using a statistical model, known as a Thurstone score.

The Thurstone scores were then used as inputs in the social welfare function (SWF). The SWF assumes that preferences follow various standard principles of rational choice (in particular, transitivity and the Pareto principle).

The first set of questions established the inequality aversion parameter of the SWF by asking questions where the groups differ in relation to their life expectancies. The second set of questions onwards elicited preferences the timing of ill health (childhood v adulthood) and the severity of ill health (25% v 100%), which together with life expectancy, make up lifetime health prospects. The additional study explored using a different level of severity (50%), and different ages (10-18 and 18-40).

In the third set of questions onwards, the two groups differed in terms of responsibility and then in terms of the rarity of the condition. The final set of questions tested for the effects of specific labels (e.g. obesity).

Equity weights results

There were 582 interviews in the main study and 130 in an additional study designed to provide more data for generating equity weights. The analysis was conducted on 559 and 129 respondents, respectively.

Overall, respondents expressed preferences that diverge from QALY maximisation. It is difficult to present the results in a summary because the weightings attached to the attributes considered here change together but here are some general findings:

1. There is a general aversion to inequality. To illustrate, a marginal health benefit to a group with a life expectancy of 75 is worth about 30% more than the same benefit to a group with a life expectancy of 82.

2. There are concerns about the timing of ill health: a marginal benefit to children is worth about 80% more than the same benefit to adults.

3. There are preferences for severity: the first 25% on the individual health scale is valued as 27% of the scale in terms of its social value; the next 25% (i.e. from 25% to 50%) is worth, in social value terms, 35% of the scale. This means that the move from 50% to 100% in individual terms is worth 38% in social value terms.

4. There is a general preference to take account of responsibility: if ‘bad luck’ is given a weight of one, then ‘some bad choices’ is given a weight of 0.92 whilst ‘NHS causes’ is given a weight of 1.23.

5. Labelling has some effect with obesity-related conditions receiving less weight than some bad choices: about 0.80 as compared to 0.92.

6. There is a slight preference for condition rarity: a ‘extremely rare’ condition is given 20% more weight than a ‘slightly more common’ condition

From various analyses of the effect of background characteristics, the results do not appear to be unduly affected by the characteristics of the respondent that have been found to affect the individual valuation of health e.g. gender, age and education.

Conclusion

This project has met its aims but there is of course much more research that can and should be conducted into public preferences relating to the societal value of health gains to different beneficiaries and we should like to provide additional data to populate our general SWF approach. In the meantime, the results presented here can be used as an input into important decisions about how to equity weight QALYs.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND

It is widely accepted that one of the principal objectives of government expenditure on health care is to generate health. Since health is a function of both length of life and quality of life, the quality-adjusted life-year (QALY) has been developed in an attempt to combine the value of these two into a single index number (Dolan, 2000). QALYs are increasingly being used in the evaluation of health care interventions and have been recommended by the National Institute for Health and Clinical Excellence (NICE) for use in cost-utility analyses of health technologies (www.nice.org).

There are, however, many concerns about the appropriateness of using QALYs to inform resource allocation decisions. These relate to issues about the extent to which QALYs adequately capture the individual benefits and social value from health care interventions. In what follows, we will assume that QALYs are an appropriate currency in which to express individual health benefits. At the social level, the de facto standard in economic evaluations is that “a QALY is a QALY is a QALY” i.e. that a given health benefit is valued equally regardless of who gets it. However, there is now good evidence (albeit often from small convenience samples) that this assumption is not a good representation of societal preferences.

In relation to characteristics across which health is expected to vary, those in more severe conditions generally get higher priority (Nord 1995; Dolan, 1998; Ubel, 1999). However, two UK studies suggest that people may wish to give benefits to those in better health if those in poorer health are left in states considered to be below a threshold (Roberts et al, 1997; Dolan and Green, 1998), and a recent study from Denmark suggested that a concern for severity might depend on the particular health dimensions affected (Gyrd-Hansen, 2004). Much of the evidence is supportive of giving greater weight to younger people (Nord et al, 1996; Rodriguez and Pinto, 2000; Tsuchiya, 2001). However, only one study has attempted to control for the size of the benefit, so it is not possible at this stage to say how a given benefit is weighted across different age groups (Tsuchiya et al, 2003). Some studies have also shown a willingness to target the lower social classes (Lindholm and Rosen, 1998; Dolan et al, 1999; Emmelin et al, 1999; Dolan and Tsuchiya, 2002; Abásolo and Tsuchiya, 2003).

In relation to characteristics across which health itself may not vary, there is a consensus that we should discriminate in favour of those with dependants (Neuberger et al, 1998; Dolan et al, 1999; Browning and Thomas, 2001), although Edwards et al (1999) is an exception. There is some evidence to support the view that people who are themselves considered to be responsible for the ill health should be given lower priority (Ubel et al, 2001; Dolan and Tsuchiya, 2002). However, there is no evidence on whether higher priority should be given when somebody other than the patient (such as the health service) is considered to be responsible for the ill health. That said, there is limited evidence that people are willing to give priority to second-time liver transplant patients, where the first transplant may have been interpreted as a medically caused failure (Ratcliffe, 2000).

Therefore, the evidence currently available suggests that simply maximising QALY gains across the population is not the best use of resources according to people’s preferences. However, it does not allow us to estimate equity weights for QALYs. This is the purpose of this project. The specific aims are now set out in a little more detail.

1.2 AIMS OF THE PROJECT

There are three main phases to this project, related to three principal aims:

1.2.1 To determine the relevant attributes/levels over which to elicit preferences

The evidence referred to above provides some guidance about the attributes that the general public might wish to take into account when determining the relative societal value of health gains to different beneficiaries. However, we need to consider the relevance of these attributes in the context of macro level policy decisions, since some of the existing evidence relates to small groups or 'bedside' decisions. Moreover, the evidence provides very little guidance about the levels of categorisations of those attributes that are relevant to people e.g. age may matter when the choice is between a child and an adult but not when it is a choice between a 40 year-old and a 50 year-old. Phase 1 of the project addresses this aim and is reported in Chapter 2.

1.2.2 To develop methods which facilitate the elicitation of stable preferences

There would seem to be general agreement across economics, psychology and political science that, if people's preferences are to be used to help guide policy, then those preferences must be stable. We are more likely to elicit a stable set of preferences if we first elicit general beliefs and attitudes. General political beliefs and attitudes regarding the roles and responsibilities of the NHS may also be drivers of preferences for particular distributions of health outcomes over other people. By eliciting such views, we may serve to better explain our responses i.e. just not in terms of standard socio-economic characteristics but also in terms of people's beliefs and attitudes. In addition to this, we consider the degree to which more stable preferences might be elicited through a discussion group prior to individual interviews. Phase 2 of the project addressed this aim and is reported in Chapter 4.

1.2.3 To generate 'equity weights' from the preferences of the general population

Our specific goal here is to elicit preferences from members of the general public that can be fed into a social welfare function (SWF), which allows us to estimate the weight given to the health of one population group relative to another. Most of the studies in this area have adopted a social perspective, in which respondents are asked to consider allocation decisions that they personally may not be affected by. We propose to do the same. Standard economic theory is concerned with the optimising behaviour of self-interested individuals, and so many economists are rather distrustful of preferences that contain no self-interest at all. However, although self-interest exists, it does not necessarily follow that it must – or should – be the basis for social welfare evaluation (Menzel, 1999; Dolan et al, 2003).

From an empirical perspective, even the experimental studies most cited as evidence of in-group favouritism also show that self-interest is constrained by considerations of fairness (Tajfel, 1982). Indeed, there is a large psychological literature showing that individuals will frequently forego personal advantage in favour of distributions that they regard as more equitable (Lerner, 1980). All in all, we contend that the citizen perspective is an appropriate one and one that is best suited to providing the kind of data required for equity weights for QALYs. The results from the main study to elicit equity weights (Phase 3 of the project) can be found in Chapter 5. We now provide a little more detail about the three phases of the project.

1.3 OVERVIEW OF THE PROJECT

Phase 1 is a series of preparatory studies, which aims to identify the set of attributes that are examined in the project, and how their levels are set. It also explored the issue of preferences over concentration and dispersion of health benefits across a given population. Phase 2 of the project is a methodological study to examine the impact of deliberation on public preferences, thus informing the design of the main survey carried out in Phase 3.

1.3.1 Phase 1: what to elicit preferences over

Phase 1, reported in Chapter 2, consisted of four studies. Studies A and B used focus groups with members of the general public to identify the relevant attributes and to identify meaningful levels of these attributes, respectively. Study A was a qualitative study involving 15 members of the public in group sessions. The attributes presented for consideration were: age, social class, length of time with condition, dependents, quality of life without treatment, and whether the condition was caused by NHS negligence. All of these attributes were considered relevant in Study A in that respondents chose to depart from simple QALY maximisation in order to take them into account.

Study B was a more quantitative exercise, involving 42 members of the public in group sessions. The study was carried out in two separate rounds, which varied the number of QALYs that would have to be sacrificed to target a priority group (100 QALYs across 1000 individuals versus 500 QALYs across 1000 individuals). As expected, when the opportunity cost of diverging from QALY maximisation was larger, a smaller proportion of respondents were ready to do so. Attributes and levels for which respondents were willing to sacrifice health gains across both levels of QALY sacrifice included quality of life without treatment (when the lower level was 40%), time with condition (where one party had the condition for one year or more) and age (15-24 year-olds relative to 65-74 year-olds; and under-5s relative to 25-34 year-olds).

The aim of study C was to identify the most policy relevant combinations of attributes from a postal survey of NHS staff (n=172). The most common result across all attributes was that the attribute is relevant to NHS policy “depending on what else is known”. No single attribute was regarded by a majority as relevant to NHS policy “no matter what”. The three attributes where the majority of respondents thought the attribute in question “would be relevant depending on what else is known” were: quality of life without treatment (40% as opposed to 70%), length of time with condition (more than one year as opposed to recent) and age (5-25 as opposed to 60-80). The next most important attribute, but not a majority view, was NHS negligence.

Based on the results of studies A, B and C, three attributes were recommended for taking forward into later phases of the project: age, quality of life without treatment and responsibility. Age was clearly important but there was no real consensus about when age mattered for policy and when it did not – except in the case of children versus adults. This is the focus of our comparisons. For quality of life without treatment, a value of 25% was chosen for comparison with full health. We can be confident that 25% health meets the threshold for the general public to treat quality of life as significantly lower than full (100%) health. In terms of responsibility, NHS negligence appears to be somewhat important. Since it seems unrealistic to contrast NHS responsibility with all other causes, or with 100% patient responsibility, it was decided to present this attribute as three categories: NHS responsibility and no patient responsibility; no NHS responsibility and limited patient responsibility; and no NHS responsibility and no patient responsibility.

Study D explored preferences relating to the concentration and dispersion of benefits across beneficiaries who are equal in all relevant aspects. Whilst people may prefer to spread out health benefits to a larger number of patients than to concentrate on a smaller number if the benefit per person is large enough, they may also prefer to concentrate than to disperse if the benefits per person from dispersion are below a certain threshold. From 68 respondents surveyed in group meetings, 2.6 years was identified as the threshold or ‘tipping point’. This finding is used to guide the design of trade-off exercises used in the remainder of the project i.e. we make sure that the difference between two groups in any one scenario is at least 2.6 years.

1.3.2 Phase 2: how to elicit stable preferences

Phase 2, reported in Chapter 4, surveyed members of the public using two different designs to determine which of them was more conducive towards eliciting stable preferences. The “resource Intensive” (RI) design involved a group discussion with fellow participants prior to an individual, face-to-face interview and the “interview only” (IO) design did not involve the group discussion stage. The objective of Phase 2 was to identify the impact on peoples’ preferences of the group discussion, and of the opportunity to deliberate over the issues between this group discussion and the individual interview. The stability of preferences was captured by administering a series of attitudinal questions on resource prioritisation at multiple time points in the study design. There were 56 respondents in the RI design and 232 for the IO design.

The results suggest that design appears to have had no significant effect on the willingness to prioritise different groups. Deliberation does have an effect on general prioritisation preferences, but this appears to be limited to the different stages within the RI design rather than between the two designs. However, given the sample size, and the different distributions of background characteristics across the two sub-samples of this study, the interpretation of this is not quite so straightforward. At a practical level, the IO design proved to be far more practical in terms of recruiting respondents and it was much less costly. Since the prior elicitation of beliefs and attitudes appear to be more important in generating stable preferences than discussion groups, we use the IO design in Phase 3.

1.3.3 Phase 3: generating equity weights

The main preference elicitation task designed to elicit equity weights consisted of six “choice sets”, each comprising four pairwise choices, where the respondent was asked to choose between two scenarios, both involving two population sub-groups. For example, the first scenario may represent relatively high average health across the two groups but with a relatively high health inequality between them, whereas the second scenario may represent relatively low average health across the two groups but with a relatively low health inequality between them. The respondent was asked to indicate their preference between the first and the second scenarios.

Each choice set was carefully designed to enable the elicitation of particular preferences (details are reported in Chapter 3). The first choice set established the inequality aversion parameter of the SWF by asking question relating to two groups that differ only in relation to their lifetime health prospects. The second choice set onwards involved poor quality of life (25% health), during childhood (<18 years old) or during adulthood. In the third choice set onwards, the two population subgroups differed in their responsibility for their health. In choice sets three and four, they had different responsibility status, where these are described using fairly general and abstract phrases, whereas in the fifth choice set, these were described using more specific labels. In the sixth choice set, the population subgroups are afflicted by diseases that have contrasting levels of rarity (this attribute was added after discussions with NICE). An additional study explored using a different level of severity (50%), and different ages (e.g. 10-18 and 18-40).

A total of 582 members of the public were interviewed in the main study, and a further 130 in the additional study. The main findings of the analysis of Phase 2 and Phase 3 data were that, on average, people have preferences that diverge from simple unweighted QALY maximisation. They prefer to trade-off overall health for a more equal distribution, they regard 25% quality of life as having more value than a quarter of full health (but this is not significant), they give more value to a year of life in childhood than at adulthood, they give higher priority to ill health caused by the NHS; and finally, they give higher priority to rare conditions than common conditions (but this is not significant).

In what follows, Chapter 2 reports on Phase 1 of the project, where the attributes and their levels were selected. Chapter 3 presents the SWF approach used in this project, and the design of the preference elicitation tasks. Chapter 4 reports on Phase 2 of the project, where the study design was chosen. The results of the main preference elicitation tasks, alongside the sensitivity analyses and additional study, are presented in Chapter 5. Chapter 6 provides an overall discussion.

CHAPTER 2: DETERMINING THE ATTRIBUTES AND LEVELS

This chapter reports on the four preparatory studies in Phase 1 of the project. These studies were designed to select the attributes, and the levels of those attributes, that the general public would distinguish between in later phases. The first study was designed to identify the relevant attributes from discussion groups with members of the general public. The second study focussed on the general public's views about the relevance of different levels or categorisations of the attributes. The third study elicited the views of decision-makers about the relevance of various attributes in policy-making. The fourth study was designed to find out how small the health benefits for each individual had to become for members of the general public to switch from generally wanting to disperse overall health benefits widely to wanting to concentrate them in fewer people. This 'tipping point' is directly relevant to the interpretation of the results in Phases 2 and 3.

2.1 STUDY A: GENERAL PUBLIC VIEWS ABOUT ATTRIBUTES

2.1.1 Methods

2.1.1.1 Focus groups

Focus groups have been designed to elicit opinion-type responses on a defined topic of interest (Stewart and Shamdasani, 1990) and were considered appropriate for use here. Focus groups involve small group discussions designed to understand respondents' views and also why they hold those views from their own perspective and in their own words (Kvale, 1996).

Each group discussion began with general descriptive information about the nature of scarcity in health care, the need for prioritisation at a macro level, and an explanation of QALYs. An interview guide approach was utilised, where lists of questions were covered, but informality was retained. This flexibility enabled the researcher to enter novel areas and tailor the questions according to the respondents' positions, thereby producing rich data (Banister et al, 1994).

Respondents were asked to choose between two groups of patients (Group A or Group B) on the basis of who they would prefer to give additional NHS treatment to. The two groups consisted of equal numbers of patients (500 patients in each), both received 10 QALYs per patient with standard treatment, and the cost of the new treatment was the same. Those in Group A would benefit by 1000 QALYs from additional treatment and those in Group B would benefit by 900 QALYs, so that a simple QALY maximisation principle would give the additional treatment to Group A. Respondents were asked whether they were happy with this QALY maximising solution and to explain the reasons behind their responses.

There was a warm-up example, followed by eight questions. Each question was represented on a flip chart and placed in front of the respondents. The combinations of patient attribute used were as follows and draws on those attributes that have been found to be important in the literature to date (see Dolan et al, 2005). Gender (number 0) was used as the warm-up exercise.

	500 people in Group A who will get 1000 QALYs with additional treatment	500 people in Group B who will get 900 QALYs with treatment
0	Male patients	Female patients
1	Patients aged 65 to 80	Patients aged 25 to 40
2	Patients aged 5 to 15	Patients aged 25 to 40
3	Patients in social class 1	Patients in social class 5
4	Patients in social class 5	Patients in social class 1
5	Patients who have had the condition for 2 years	Patients who have had the condition for 10 years
6	Patients who do not have dependents	Patients who have dependents
7	Patients with a QOL at 70%	Patients with a QOL at 30%
8	Patients whose condition was caused naturally	Patients whose condition was caused by NHS negligence

Note that the selection of the attributes at this stage is based on what is prevalent in the empirical literature where members of public were surveyed for their views over priority setting, including micro level priority-setting. In other words, the selection does not necessarily reflect what may be regarded as appropriate in the particular context of macro level NICE-type decision-making.

Respondents were informed beforehand that the replies and arguments that they provided would not affect the treatment that they would receive from the NHS, and emphasis was placed on the anonymity of individual responses to encourage openness. There was also an opportunity at the end of the focus group sessions for respondents to voice any issues that they felt may have been missed out. This helped ensure that a full picture was gained from their perspective.

The focus groups typically lasted between 60 and 90 minutes with a time schedule being agreed upon with the respondents prior to the discussions. All focus groups were audio taped with the respondents' consent and transcribed in full to ensure that the data were preserved in 'raw form' helping to remove the effect of the researchers throughout the data collection (Seale, 2000). Notes were also taken by one researcher.

Once the focus groups were completed, respondents were asked to complete a background questionnaire. Each respondent indicated their gender, age (in years), whether they considered themselves to be in social class 1, whether they considered themselves to be in social class 5, whether they considered themselves to be part of an ethnic minority group, their highest educational qualification, how many children they had, how many children they had under 18 years of age, their QOL (0% - 100%).

They were asked their thoughts regarding the content of the focus group discussions, by selecting as many items as they thought applicable from the following list: Very Unrealistic; Insightful; Interesting; Too Complicated; Boring; Enjoyable; Informative; Too Sensitive a Topic; Too Quick; Useless. This same set of items were used in the other general population samples in Phase 1, Respondents were then thanked for their participation and given their payment of £15 for taking part.

2.1.1.2 Recruitment and respondents

Respondents were approached in Sheffield city centre, given a description of the study and advised they were under no obligation to participate in the study. Potential respondents were informed that by participating they would receive a payment of £15 as a token of appreciation. A recruitment questionnaire was given to those who agreed to take part in the study to ensure that a wide variety of respondents were selected for the focus groups. The recruitment questionnaire asked respondents to indicate their gender, age (in years), employment status, whether they continued education after the minimum school leaving age, whether respondents had a degree or equivalent, whether the respondent, someone in their family, or someone who they cared for had experienced any serious illness in the past, whether they had any children and whether any of their children were under 18 years of age.

Respondents received a study pack in the post containing a cover letter, a participant information sheet briefly describing the study and their role in the study, a map of where the focus groups were to be held, and a consent form. Respondents handed in the consent form at the beginning of the focus group sessions. We intended to conduct four groups with about five people in each: it is likely that the discussion themes will be exhausted after four sets of discussion. We sought to sample people from a range of backgrounds so as to enhance the general representativeness of the results.

Out of the 25 individuals that agreed to attend the focus group session, 60% (15 respondents) actually took part in the study. Respondents were placed into one of four groups consisting of three, three, eight and two respondents each, held in seminar rooms at the University of Sheffield. The groups consisted of seven women and eight men ranging between the ages of 23 to 64 (M = 35.5). Seven of the respondents considered themselves to be in social class 5, and 1 in social class 1. The perceptions of the respondents' QOL ranged between 10% and 100% (Mean = 68.36, Median = 75.00, SD = 28.07). The five most frequently chosen 'thoughts' of the focus group sessions include: interesting (N = 12); insightful (N = 4); informative (N = 4); enjoyable (N = 3); and too quick (N = 2).

The ten respondents that did not attend the session consisted of six females and four males, five aged between 18-25, one aged between 26-35, two aged between 36-45 and two aged between 46-55. In terms of experience with ill-health, the non-respondents included three who cared for others, two who had experienced serious illness themselves, and a further two who had experienced some other serious illness in their family. In employment terms, the non-participants included four students, four homemakers, one retiree, and one employed person. Six of the non-respondents also had children, of whom five currently had children aged under 18.

2.1.1.3 Analysis

The purpose of the thematic analysis used to examine the data was to explore the issues important to the general public when prioritising health benefits based on patient attributes. Once the data were transcribed verbatim, repeated reading allowed the researcher to become familiar with the data. Data were independently coded into units of categories which were generated into themes from the coded units. A theme, as used in this study, is "an abstract entity that brings meaning and identity to a recurrent experience and its variant manifestations. As such, a theme captures and unifies the nature or basis of the experiences into a meaningful whole" (DeSantis & Ugarriza, 2000).

The focus group findings are presented below in terms of these qualitative themes that describe the framework with which respondents evaluated the various groups across which prioritisation decisions could be made. In order to ensure the coding did not fix meaning too early in the analysis, initial coding acted as signposts or indexes to interesting pieces of data. Later coding was applied when the researcher believed that the index words had regularity or stability in the way in which they occurred across different contexts, helping to increase internal reliability (Seale, 2000).

2.1.2 Results

A number of themes emerged from the focus group data, and these are presented below. This thematic analysis served to integrate these findings into our understanding of why people preferred to treat certain groups over others. Each of the themes is presented in terms of the attributes (age social class, length of time with the illness, the presence of dependents, QOL and NHS negligence).

2.1.2.1 Social class

2.1.2.1.1 Money / alternative treatment

Money / alternative treatment emerged as the most frequently occurring theme. Many of the respondents associated higher social classes with higher income brackets and thought that these patients could afford to purchase private health care, and therefore preferred to treat lower social classes (i.e. those that couldn't afford health care outside the NHS). Several respondents' comments represent the strength of these views:

> "They can buy their own, they are essentially higher paid so they can pay for it themselves"
>
> "When it comes to social class they can afford to pay private, why can't the national health give it to Group B [social class 5]?"
>
> "If people who could afford it go elsewhere and leave the resources for the people that can less afford them"
>
> "They [social class 5] can't afford the alternatives"

2.1.2.1.2 Awareness of health issues

Many of the respondents perceived that those in lower social classes are less aware of the issues surrounding their health and have less access to the information and therefore many preferred to treat those in social class 5:

> "I might be tempted to go for Group B on the basis that they are in social class 5 are not aware or possibly well read and don't know the standard health issues as much as people in social class 2 [sic.]"
>
> "I would give 1000 to social class 5 because it has already been recognized that people in social 1 … are more aware of their personal health risks"

One participant did express opposing views when she explained that all social classes have access to information available to improve their own health:

> "Everyone gets the same information via the media"

> “I think I’d stick with Group A [social class 1, QALY maximisation] again, social 5 can still read the papers get the same information if they are concerned about their health, just as much as social class 2 [sic.]”

However others indicated that even if social class 5 did have access to information it was highly unlikely that they would fully understand it:

> “Can they [social class 5] take it on board, do they actually understand a lot of what’s in the press and do they buy their own newspaper?”

2.1.2.1.3 Lifestyle

Another theme that was regarded as important for treating lower social classes was lifestyle. Over the years numerous researches have illustrated that the general public tend to prefer treating those who look after their health, rather than those that don’t. However despite these advances, when respondents believed that the negative health behaviours of patients in lower classes were the result of class position, many preferred to prioritise social class 5 as it was thought that they were at a disadvantage:

> ‘You could argue there’s loads of points [reasons to give the health benefit to people in different social classes] we just go to the barebones of society…..Sort of social class 5 you imagine people in the back streets of Glasgow eating deep fried mars bars and smoke 40 fags a day’

> ‘I would give it to social class 5 because it has already been recognized that people in social class 1 …..smoke less, drink less, play squash and active sports such as that, taking more care of their health, their life span is naturally longer. So I would give 1000 to social class 5’.

2.1.2.2 Dependents

2.1.2.2.1 Extra benefits and costs

The analysis illustrates that the extra benefits and costs arising from treating/not treating patients with dependents was mentioned over and over by the respondents. They believed that treating people with dependents would result in more QALYs overall as many felt that the dependents would benefit too from the treatment via the increased health of the carer. Likewise respondents believed that if patients with dependents were not given the additional treatment the dependents would suffer as a result of the carer’s ill health. The importance of this issue was a consistent theme echoed by the respondents:

> “people in Group B [those with dependents] would be better off than they were and hopefully that will filter down to their children, who will benefit as well”

> “Yes I’m strongly influenced that way as well, Yes I think Group B [people with dependents] the better treatment because you’d feel that it is going to benefit more people overall because it would benefit their children”

> "You wouldn't want to think there were children suffering because their parents were suffering"
>
> "I would agree with that, the welfare of the child has got to be taken into consideration first"

2.1.2.2.2 Carers

Respondents were also particularly concerned about leaving the dependents without a carer if the illness was life threatening or meant that the carer was simply too sick to look after the dependent themselves:

> "Well you have got to take the welfare of the child into account, single parent, young child, is there someone to look after the child, parents not around that would sway me in favour of Group B [patients with dependents]"
>
> "Give it to the ones with dependent children. Cos they have got to look after children"
>
> "Have the people with dependent children got partners to look after them through illness, as opposed to the people with no dependent children having no one to look after them. Are we looking at single parent families, nuclear families, or are the people with no dependent children single?"
>
> "Give the 1000 QALYs to those with no dependent children and tell them to look after the people who have got dependent children" Group laughter.

2.1.2.3 Age

2.1.2.3.1 Benefits to society

The benefit of treating patients of a working age has been recognized in previous studies as an important factor treating younger patients (e.g. 25 to 40 year olds) over older patients (e.g. 65 to 80 year olds). The focus group respondents equally emphasised the necessity of treating younger patients due to the accompanying benefits for society:

> "The contrast between A [65-80 year olds] and B [25-40 year olds] even though they have less QALYs in Group B the younger people may be more beneficial to society so the younger people getting more QALYs whatever"

This was also true for one participant when asked to prioritise 5 to 15 year olds or 25 to 40 year olds:

> "The age range of Group B [25 to 40 year olds] would be working people, people who are eligible for work whereas the young ones they wouldn't, no so I agree with Group A [implying a preference to treat 25 to 40 year olds rather than 5 to 15 year olds]"

Yet many respondents also perceived many societal benefits by treating 5 to 15 year olds over patients aged between 25 and 40. This was due to concerns of preserving the future of society:

> "Yes, but to give it to younger group you're looking more for the community and future init, they're future t'young people so"
>
> "They're getting, they're our future aren't they"

2.1.2.3.2 Fair innings

In addition to the benefits for society described above, the focus group respondents also described that fair-innings contributed to their preferences of treating younger patients. They stressed that as older patients had seen most of their lives it was fairer to give the additional treatment to younger patients that hadn't:

> "Age 65 – 80, it sounds cruel, but they have had a reasonable innings, I'd be more inclined to give that quality of life to Group B [25-40 year olds]"
>
> "You would be better off to give the treatment the quality of life to the 25 to 40 year olds than the 65 to 80's, they've had a life"
>
> "Think most people prefer to think that children should receive better treatment than people who had seen most of their life"
>
> "5 – 15, you've got more things to look forward to and that you know what I mean"

2.1.2.3.3 Entitlement

Conversely as many of the older patients were perceived to have made numerous contributions to society such as paying for national insurance etc, some focus group respondents felt these groups of patients should be entitled to the additional treatment more so than younger groups (those aged between 25 and 40):

> "I know where you're coming from 'cos you've worked all your life, and you've paid all this tax and national insurance, you deserve to get something back from what you've paid"
>
> "I would say now that age 60 to 80, 65 to 80 would need more because of the contribution that they have made during their lifetime"

2.1.2.3.4 Emotional attachment

When asked to prioritise very young patients (5 to 15 year olds) or 25 to 40 year olds, many respondents, especially those with children, acknowledged an emotional attachment for children, and therefore preferred to give the health benefit to the very young:

> "I've got to admit that looking at the boxes there is a bit more relief [5 to 15 year olds receiving more QALYs from additional treatment rather than older age groups], I don't know why but I haven't thought about this question before but immediately I was sort of a bit more relieved that it was that age group [5 to 15 years, QALY maximisation] rather than that age group [25 to 40 year olds, QALY maximisation]
>
> "I completely understand that the emotion of being that way [preferring to treat children], I'm very much relieved"
>
> "I think I'm going to stay out of this one as a mother you tend to go for children"
>
> "Yes but I was thinking you were saying that is an issue with the very young age group, people get more emotional, I got more emotional if QALY's were lower for younger age group. If the QALY's were lower I don't know how I would react. I don't know how I would react if the QALYs were much lower for the younger age group"
>
> "Yes, if you've got kids your kids mean everything to you don't they"

2.1.2.3.5 Intrinsic values

It was also apparent that some placed an intrinsically higher value on a child's life, and therefore choose to treat the younger respondents over those aged 25-40:

> 'We always put children on the highest pedestal, don't we'
>
> "I think exactly the same [treat the younger age group], because they are young"
>
> "You are always going to take the younger generation aren't you"

2.1.2.4 Quality of life (QOL)

2.1.2.4.1 Acceptable levels

The importance of increasing patients' low QOL to an acceptable level was seen as a recurring theme. Many felt that a 70% health state was already a decent health state and that a 30% quality of life was too poor for patients to remain in:

> "yeah, because 70% is a good percentage anyway, you better off getting the 30% further up"
>
> "I think most people would argue that you should give slightly lesser medical care 900 QALYs to Group B [lower QOL, 30%] and up their quality of life yeah and the people in Group A [higher QOL, 70%] are still getting to be on 70% quality of life which is what you would get group b towards that kind of level and that would be the best benefit"
>
> "No it doesn't matter if somebody's quality of life is that low already [30%] you can drag that up, whereas people in Group B [70% QOL] are considerably better off than people in Group B [30% QOL] anyway"
>
> "70 is not it's not that bad out of 100 is it, 30 is a lot worse than 70"

2.1.2.4.2 Value of change

Within the focus groups, respondents tended to place a greater value on a change from a very low health state to a reasonable health state more than a change from a reasonable to a perfect health state:

> "Like the number of years they have been waiting, they would have a better improvement because of that"
>
> "So we all say that the change from 30 to 60 is worth more than a change from 70 to 100?"
>
> "Yes"
>
> "Yes, because their quality of life is so low"

2.1.2.4.3 Deservedness

It was also perceived by many that those in a lower QOL were more deserving of the treatment than those in a higher QOL:

> "I say give the 900 to the people with the 30% QOL because if they are disabled, possibly severely disabled and they deserve some kind of support for the life they have got"

> “Cos even though they are not getting as many QALYs they, they deserve you know it would help to”

2.1.2.5 Length of time with the illness

2.1.2.5.1 Length of suffering

Within this attribute, the length of time that patients were suffering for was the most dominant theme. The focus group respondents felt that those patients who had suffered with the illness for 10 years had suffered long enough and therefore should not have to endure the illness for any longer:

> “Whereas them what have had it for 10 years I suppose you could say they have suffered enough do you know what I mean”
>
> “They've waited that long do you know what I mean they shouldn’t have to wait any longer”
>
> “So if you are going to alleviate their condition out of it having suffered longer, the stuff to ponder, even though they are not that group aren't quite getting the benefits of what you might be able to afford, that seems a more compassionate way forward”
>
> “I would say give 900 QALYS to people who have had the condition for 10 years, because they have been suffering a long time”

2.1.2.5.2 Outcomes of the treatment

In contrast to the above preference, some felt that the outcomes of the treatment would not be as good for those that had been waiting longer for the treatment. They perceived the condition to be worse when it was present for longer and that the benefits of the additional treatment would not be as great for those that had had the condition for longer:

> “Depending on what the condition is, if after 2 years they have a better chance of making a full recovery it’s better than somebody who had the condition for 10 years with maybe less realistic prospects. On balance I would say give 1000 for those with the condition for 2 years”
>
> “There should be better results, rather than if they’ve had it for 10 years”
>
> “They have got a better chance of like catching it at 2 years whatever the illness is”
>
> “They’ve had it they've had that condition for 10 years, and they’ve only had it for 2 years so, it’s better to give them (those that have had the condition for 2 years) the quality because you might be obviously you might be able to cure whatever disease they have. They’ve had it for 10 years now, so..”
>
> “If you go for Group A [those that have had the condition for 2 years] it would be totally improving the life of those people who have just had the condition for two years”
>
> “It might not be curable whatever it is, [those that have had the condition for 10 years]”

2.1.2.5.3 Adaptability

The final theme that emerged was the adaptability of the illnesses. Focus group respondents felt that those that had had the condition for longer would be more adaptable to the illness than patients that had had the illness for less time and therefore some decided to give the treatment to those that had had the condition for 2 years:

> "With its been 8 years they could know a bit about it, so it's better to treat them who've had it for 2 years rather than, if they've had a condition for 10 years then them have had it for 2 years, they should be 8 years knowledgeable about whatever it is"
>
> "You can live with it a bit longer can't you"
>
> "It's not gonna hurt for them to just to you know [live with it a bit longer]"

2.1.2.6 NHS negligence

2.1.2.6.1 Caused illness

Many felt that if the illness was inflicted on the patient, they preferred to give treatment to the NHS negligence group rather than the natural causes group:

> "It's not natural is it, it's not a natural illness; it's something that somebody else has caused, so it should be treated"
>
> "Yeah but it's still gonna go with the negligence every time, because it's not natural"
>
> "It's a man-made mistake"
>
> "It's been inflicted rather than"

2.1.2.6.2 Rectify mistakes

The final theme that emerged from the analysis was 'rectify mistakes'. Some felt the patients that had been neglected by the NHS should receive the additional treatment as the NHS had a duty to rectify their mistakes:

> "You really expect it to be sort of they should put it right, I perceive that their situation is put right"
>
> "Give 900 to the people caused by NHS negligence and restore faith in the NHS to do its job"
>
> "I think if I caught a bug in hospital I would expect the NHS to put it right"
>
> "Let NHS rectify mistakes it's made"

2.1.3 Conclusion

Findings from the focus group interviews indicate that the relative societal value of health gains to different beneficiaries varies on personal attributes. Specifically, the following attributes attracted higher priority: those of a lower social class, patients with dependents, younger age, lower quality of life, longer length of time with illness, and illness caused by NHS negligence. When respondents were presented with a choice to either QALY maximise (1000 QALY gain) or to prioritise the disadvantaged groups (900 QALY gain), the majority choose the latter option.

'Money/alternative treatment' 'awareness of health issues' and 'lifestyle' were themes that emerged for prioritising lower social classes. Patient groups with dependents were also preferred due to the 'extra benefits and costs' arising from treating/not treating these groups and because many did not want to leave the dependent without a ‘carer’ if the illness deteriorated. The perceived societal gains from prioritising younger patients meant that many advocated a preference to treat those of a younger age. Respondents also prioritised younger people on the basis of the ‘fair innings’ argument. Interestingly, personal factors also played an important role in the personal preferences on individuals. A large majority, especially those with children, felt emotionally attached to younger patients (5-15 year olds) and tended to place an intrinsically higher value on a child’s life. On the other hand, some thought that older patients were entitled to the additional treatment, due to the contributions (e.g. national insurance) made throughout their lives.

When faced with a choice between treating lower (30%) quality of life over higher (70%) quality of life patients, three themes emerged. First ‘acceptable levels’: a 30% QOL was perceived as an overly ill state to tolerate so many felt that every effort should be made to increase the health state of these patients. Second ‘value of change’: a relatively higher value was placed on increasing poorer health. Third, ‘deservedness’: those in poorer health deserved the treatment more. Respondents also expressed the view that the length of time with the condition was an important attribute. The majority advocated that the NHS should prioritise those patients that had suffered with the illness for a longer period of time. Finally, respondents seemed to be in favour of allocating additional health benefits to illnesses resulting from NHS negligence over and above those that have been caused by chance. Some indicated that if the illness was imposed by the NHS, then the NHS had a duty to rectify its mistake.

Taken together, these results illustrate that the relative societal value of health gains to different beneficiaries varies depending on the personal attributes of the patients. The results are broadly in line with those found elsewhere in the literature (see Chapter 1 and Dolan et al, 2005) and suggest that these preferences have a reasonably clear and coherent basis. The results from this study were fed into the next preparatory study, where different levels of these attributes will be determined.

2.2 STUDY B: PUBLIC VIEWS ABOUT LEVELS OF ATTRIBUTES

As with Study A, the sessions began with an introduction to scarcity in health care, the need for prioritisation at a macro level, and an explanation regarding QALYs. Respondents were then given the questionnaire consisting of 12 questions, which were processed in an interactive manner. The facilitator first explained how the questionnaire works, and went through the first example question together (this question is reproduced as Appendix 1). Respondents were then asked to look at the first question. The facilitator explained the key attribute in the question, followed by a brief discussion of relevant issues, and then respondents completed the questionnaire on their own. Once all respondents had completed the question, the whole group moved to the next question, and the same steps were repeated.

Each question was a series of pairwise choices between patient groups A and B similar to those in Study A (e.g. 70% and 50% quality of life). Group A always received more QALYs (1000 QALYs), and, based on the results of the initial study, more preferred attributes were assigned to Group B with less QALY gains (900

QALYs). Respondents were directed to the next pairwise choice depending on their response to the previous pair. Group A in the second pairwise choice was the same as that in the first pair (i.e. 70% quality of life), but the level of the attribute in question was changed for Group B (if a respondent had chosen Group A at the first pair, then Group B a the second pair was 40%; if a respondent had chosen Group B at the first pair then Group B at the second pair was 60%).

There were three pairwise choices per question irrespective of the pathway taken. This eliminated the incentive to choose the patient group that led to fewer numbers of pairwise choices. To reduce confusion in the task, the patient group receiving more QALYs always appeared on the left hand side of the scenario, and the same question structure was maintained throughout.

The table below summarises the initial pairwise choice in each question. Of the attributes included in Study A, NHS negligence was dropped from this part of the study, since it was judged that exploring different degrees of this concept was too complex. The warm-up example using gender had to be replaced (given that there are no degrees of gender). Since ‘social class’ seems to be affected by notions of affordability, income quartiles (“rich: the upper 25% of the population”, “comfortable: next 25% of the population”, “managing: next 25% of the population”, and “scraping by: next 25% of the population”) were introduced as the warm-up example, to allow an opportunity for respondents to express views on this attribute. The main social class question was followed by another that assumed there were no opportunities to purchase health care outside the NHS. Affordability could not enter into considerations here, and this question therefore controls for any effect from this factor.

Two questions on quality of life were used, the first aiming to identify the lower boundary (where lower quality of life is unacceptably low) and the second for the upper boundary (where lower quality of life is not meaningfully different from the reference quality of life at 70%). There were two questions on length of time with condition, one in a neutral setting and another in a waiting list setting. The duration of 2 years and 10 years was judged to be too long for the waiting list example, and was reduced to ‘recent’ and 6 months. This reduction of the duration was also in line with the finding from the initial study that some respondents clearly took the 10 year duration as something that interacts with the patients’ capacity to benefit from the additional treatment.

In response to the views in the initial study that whether the patient had a partner who could look after the child should be considered under the dependents attribute, the questions now took this into account. Also taken into account was whether the dependent was a child or other family members with a disability. Included in the questionnaire were six inconsistency measures, where a respondent will be presented with a pair that they have already answered. If a respondent contradicted their own earlier response, this was recorded as an inconsistency.

There were opportunities throughout the sessions for respondents to voice any opinions that they might have had regarding the task. Respondents were assured that their responses were confidential, and anonymity was preserved by providing respondents with unique identification numbers. Sessions lasted between 60 and 90 minutes. Sessions were again audio-taped with the respondents consent, and notes were taken throughout to aid clarity when interpreting the results.

	500 people in Group A who will get 1000 QALYs with additional treatment	500 people in Group B who will get 900 QALYs* with treatment
0	Patients who are “rich”	Patients who are “managing”
1	Patients in social class 1	Patients in social class 3n
2	Patients in social class 1 (with no health care outside NHS)	Patients in social class 3n (with no health care outside NHS)
3	Patients with a QOL at 70%	Patients with a QOL at 20%
4	Patients with a QOL at 70%	Patients with a QOL at 50%
5	Patients who got the condition recently	Patients who have had the condition for 6 months
6	Patients who are just put on a waiting list	Patients who have been on the waiting list for 6 months
7	Patients aged 65 to 74	Patients aged 25 to 34
8	Patients aged 25 to 34	Patients aged 5 to 14
9	Patients aged 75 to 84	Patients aged 65 to 74
10	Patients who are single and do not have children	Patients who are single and have 2 children
11	Patients who has a partner and 1 child	Patients who are single and have 1 child
12	Patients who are single and do not have dependents	Patients who are single and have 2 dependents

* In the first round, people in Group B were to get 900 QALYs with treatment. In the second round, people in Group B were to get 500 QALYs with treatment. Group A remained unchanged.

After respondents had completed the main questionnaire they were asked to answer a background questionnaire. Each respondent indicated their gender, age (in years), what income group they considered themselves to be in (rich, comfortable, managing or scraping by), what social class they considered themselves to be in, their highest educational qualification, how many children they had, how many children they had under 18 years of age, whether they were or had been a main carer for a disabled family member and their quality of life (0%-100%). They were also asked their thoughts regarding the content of the group sessions. Once both questionnaires were completed, respondents were thanked for their participation and given a payment of £15 for taking part in the study.

There is a concern that the difference between Group A (received 1000 QALYs with new treatment) and Group B (received 900 QALYs with new treatment) may not be large enough to warrant attention from the respondents. Therefore, a second round of groups was conducted with a separate set of participants, using exactly the same methods but this time using 500 QALYs for Group B. This means that the participants in the second round were asked to make larger sacrifices in order to diverge from QALY maximisation.

2.2.1.2 Analysis

The distributions of responses were tabulated in order to identify the preference of the median respondent. In what follows, always choosing Group A over Group B will be referred to as QALY maximising, since this is the implication. However, strictly speaking, we do not have explicit evidence to support that QALY maximisation is the reason why they made those choices.

2.2.1.3 Recruitment and respondents

The same recruitment method as before was used here. Because this study had a more quantitative focus, we sought to recruit more respondents. We sought to sample people from a range of backgrounds so as to enhance the general representativeness of the results. In Round 1, of the 45 individuals that agreed to attend, 29 respondents (64%) actually participated. They were spread out across three groups of 13, ten and seven respondents each, which were conducted in seminar rooms at the University of Sheffield.

Of the 29 respondents, 18 were females and eleven were males, ranging between the ages of 18 and 84 (M = 45.38, SD = 19.56). Of these, two of the respondents considered themselves to be in social class 1, ten in social class 2, six in social class 3 and two in social class 4. Twenty of the respondents had one or more children and four respondents indicated that they were the main carer for a disable family member. The perceptions of the respondents' quality of life ranged between 40% and 100% (Mean = 88.04, Median = 90.00, SD = 13.24). The five most frequently chosen 'thoughts' of the focus group sessions include: interesting (*N* = 21); informative (*N* = 10); insightful (*N* = 7); enjoyable (*N* = 4); and too quick (*N* = 2).

The 16 respondents that did not attend the first set of sessions consisted of eleven females and five males, seven aged between 18-25, 2 aged between 26-35, 5 aged between 36-45 and two aged between 46-55. Seven were employed, one was a homemaker and eight were students. Three indicated that they had experienced serious illness in the family and two indicated that they had experienced serious illness in caring for others. Six of the non-respondents also had children of which 5 had children aged under 18.

In Round 2 (using 500 QALYs for Group B), 72% of the respondents that agreed to take part in the study attended the session (*N* = 13), of which four were females and nine were males, ranging between the ages of 20 and 55 (M = 33.77, SD = 13.18). The results from the background questionnaire illustrated that the one participant considered themselves to be in social class 1, six in social class 2, two in social class 3 and none in social class 4. Seven respondents had one or more children, and one respondent was a main carer for a disabled family member. Quality of life perceptions ranged between 50% and 100% (Mean = 92.17, Median = 90.00, SD = 14.49). The session was held in a seminar room at the University of Sheffield. The most frequently chosen 'thoughts' of the focus group sessions include: interesting (*N* = 7); informative (*N* = 5); insightful (*N* = 4); enjoyable (*N* = 5); too quick (*N* = 1); too complicated (*N* = 1); too sensitive a topic (*N* = 1); and very unrealistic (*N* = 1).

The five respondents that did not attend the session were made of two females and three males, aged between 26 and 35 (*N* = 1), 36 and 45 (*N* = 1), 46 and 55 (*N* = 2), and 66 and 75 (*N* = 1). Three of the non-respondents were employed, one was retired and one was a student. Four indicated that they had experienced serious illness in the family; one indicated they had experienced serious illness in themselves and one indicated that they had experienced serious illness in caring for others. Three of the non-respondents also had children of which two had children aged under 18.

2.2.2 Results

The tables below illustrate the frequencies and percentages of the minimum levels of different attributes that they thought the NHS should prioritise against, and also cross tabulations between certain questions. The first two columns are for the first round of respondents and the last two columns are for the second round of respondents (n=13).

2.2.2.1 Income groups

In Round 1, when asked whether to prioritise lower income groups that gain 900 QALYs from additional treatment against the top income quartile that gain 1000 QALYs from additional treatment, slightly over a quarter of respondents support QALY maximisation, whereas the median respondent thinks that those in the second income quartile ("comfortable") or lower should be given priority relative to the top income quartile ("rich"). However, in Round 2, when the sacrifice involved was increased from 100QALYs to 500 QALYs, QALY maximisation was clearly the majority preference.

Frequencies and percentages of respondents indicating the minimum income level that they thought the NHS should give priority to when compared with patients in an income level classed as 'rich' who received the maximum amount of QALYs. (Rounds 1 and 2)

		Round 1 Frequency	Round 1 Percent	Round 2 Frequency	Round 2 percent
Valid	**"Rich"** (QALY maximisation)	7	24.1	**6**	**46.2**
	"Scraping By"	3	10.3	3	23.1
	"Managing"	0	0	0	0
	"Comfortable"	**16**	**55.2**	1	7.7
	Total	26	89.7	10	76.9
Missing	System	3	10.3	3	23.1
Total		29	100.0	13	100.0

Location of median respondent in bold.

2.2.2.2 Social class

Against a background where private health care outside the NHS is available, when respondents in Round 1 were asked to choose between social class 1, with a 1000 QALY gain, and a lower social class, with a 900 QALY gain, just under half the respondents chose to maximise QALYs, whereas the median respondent chose the group with less health gain when this group consisted of those from social class 3m or lower. However, in Round 2, QALY maximisation is again the majority preference.

Frequencies and percentages of respondents indicating the minimum social class level where alternative medical treatment is available, that they thought the NHS should give priority to when compared with patients in 'social class 1' who received the maximum amount of QALYs. (Rounds 1 and 2)

		Round 1 Frequency	Round 1 Percent	Round 2 Frequency	Round 2 percent
Valid	**Social Class 1** (QALY maximisation)	11	37.9	**5**	**38.5**
	Social Class 5	0	0	2	15.4
	Social Class 4	0	0	1	7.7
	Social Class 3 manual	**1**	**3.4**	0	0
	Social Class 3 non-manual	0	0	0	0
	Social Class 2	11	37.9	0	0
	Total	23	79.3	8	61.5
Missing	System	6	20.7	5	38.5
Total		29	100.0	13	100.0

Location of median respondent in bold.

When private health care outside the NHS is not available, similar results apply. The median respondent in Round 1 chose to diverge from QALY maximisation when Group B consisted of social class 3 manual or below, whereas the median respondent from Round 2 chose to maximise the number of QALYs gained.

Frequencies and percentages of respondents indicating the minimum social class level where alternative medical treatment is not available, that they thought the NHS should give priority to when compared with patients in 'social class 1' who received the maximum amount of QALYs. (Rounds 1 and 2)

		Round 1 Frequency	Round 1 Percent	Round 2 Frequency	Round 2 percent
Valid	**Social Class 1** (QALY maximisation)	10	34.5	**7**	**53.8**
	Social Class 5	2	6.9	2	15.4
	Social Class 4	0	0	1	7.7
	Social Class 3 manual	**5**	**17.2**	0	0
	Social Class 3 non-manual	1	3.4	0	0
	Social Class 2	7	24.1	0	0
	Total	25	86.2	10	76.9
Missing	System	4	13.8	3	23.1
Total		29	100.0	13	100.0

Location of median respondent in bold.

2.2.2.3 Quality of life

The next questions pose a choice between Group A, consisting of patients whose untreated QOL was 70% and Group B consisting of patients with a lower untreated QOL. Group A would gain a total of 1000 QALYs with treatment, whilst Group B would gain 900 QALYs with treatment. Around one-fifth of respondents in Round 1 chose consistent with QALY maximisation, whereas most respondents indicated that the worse off patients should always get priority. In Round 2, where Group B would gain 500 QALYs, the median respondent chose to prioritize those in a health state of 40% (but not those below 40%), whereas the majority of respondents chose to maximise QALYs.

Frequencies and percentages of respondents indicating the QOL level (low threshold) that they thought the NHS should give priority to when compared with patients in QOL level of 70% who received the maximum amount of QALYs. (Rounds 1 and 2)

		Round 1 Frequency	Round 1 Percent	Round 2 Frequency	Round 2 percent
Valid	70% (QALY maximisation)	6	20.7	5	38.5
	40%	1	3.4	**2**	**15.4**
	30%	0	0	0	0
	25%	1	3.4	1	7.7
	20%	3	10.3	1	7.7
	15%	0	0	1	7.7
	10%	**17**	**58.6**	0	0
	5%	0	0	1	7.7
	Total	28	96.6	11	84.6
Missing	System	1	3.4	2	15.4
Total		29	100.0	13	100.0

Location of median respondent in bold.

When the worse off patients had a higher QOL than in the previous question (ranging from 30% to 60%), less than a fifth in Round 1 continued to maximise QALYs, whereas the median respondent, alongside the majority of respondents, chose to give priority to patients with lower QOL even when the difference at baseline was 70% vs 60%. In Round 2, the median and majority of respondents chose to maximise QALYs.

Frequencies and percentages of respondents indicating the QOL level (high threshold) that they thought the NHS should give priority to when compared with patients in QOL level of 70% who received the maximum amount of QALYs. (Rounds 1 and 2)

		Round 1 Frequency	Round 1 Percent	Round 2 Frequency	Round 2 percent
Valid	**70%** (QALY maximisation)	5	17.2	**6**	**46.2**
	30%	0	0	0	0
	40%	0	0	0	0
	45%	0	0	1	7.7
	50%	2	6.9	0	0
	55%	1	3.4	0	0
	60%	**19**	**65.5**	1	7.7
	65%	0	0	3	23.1
	Total	27	93.1	11	84.6
Missing	System	2	6.9	2	15.4
Total		29	100.0	13	100.0

Location of median respondent in bold.

2.2.2.4 Time with the condition and time on waiting list

When asked to chose between those patients who have recently got the condition and who will gain 1000 QALYs from additional treatment and those patients who have had the condition for longer and who will benefit by 900 QALYs from additional treatment, 14% of respondents in Round 1 chose to maximise QALYs, whereas the median respondent, alongside the majority of respondents, chose to prioritise patients who have had the condition for longer even if it was by one month. In Round 2, where Group B will benefit by 500 QALYs from additional treatment, a third of respondents chose to maximise QALYs, whereas the median respondent chose to prioritise patients that had had the condition for one year or longer.

Frequencies and percentages of respondents indicating the length of time with condition level that they thought the NHS should give priority to when compared with patients that had recently got the condition who received the maximum amount of QALYs. (Rounds 1 and 2)

		Round 1 Frequency	Round 1 Percent	Round 2 Frequency	Round 2 percent
Valid	Recently (QALY maximisation)	4	13.8	4	30.8
	2 years	3	10.3	0	7.7
	1 year	0	0	**1**	**15.4**
	9 months	0	0	2	7.7
	6 months	0	0	0	0
	4 months	1	3.4	1	7.7
	3 months	0	0	0	0
	1 month	**21**	**72.4**	1	7.7
	Total	29	100.0	9	69.2
Missing	System	0	0	4	30.8
Total		29	100.0	13	100.0

Location of median respondent in bold.

Frequencies and percentages of respondents indicating the length of time on the waiting list level that they thought the NHS should give priority to when compared with patients that had recently got the condition who received the maximum amount of QALYs. (Rounds 1 and 2)

		Round 1 Frequency	Round 1 Percent	Round 2 Frequency	Round 2 percent
Valid	Recently (QALY maximisation)	1	3.4	3	23.1
	2 years	2	6.9	0	0
	1 year	0	0	0	0
	9 months	0	0	0	0
	6 months	1	3.4	0	0
	4 months	1	3.4	0	0
	3 months	0	0	2	15.4
	1 month	**23**	**79.3**	**6**	**46.2**
	Total	28	96.6	11	84.6
Missing	System	1	3.4	2	15.4
Total		29	100.0	13	100.0

Location of median respondent in bold.

When the same question is posed in a waiting list context, only 3% of respondents in Round 1 chose to maximise QALYs. The median respondent, alongside the majority of respondents, chose to prioritise patients who have had the condition for longer even if it was by 1 month, as in the above question. In Round 2, the median and majority of respondents chose to prioritise patients that had been on the waiting list for longer even if this was for as little as one month.

2.2.2.5 Age

When the two Groups were composed of patients of different age bands, in Round 1, where diverging from QALY maximisation entailed a sacrifice of 100 QALYs, around a quarter to a third of respondents in each question choose to maximise QALYs regardless of the combinations of patient age. In Round 2, where the sacrifice associated with diverging from QALY maximisation was 500 QALYs, around a fifth of respondents chose to maximise QALYs initially, whereas the median respondent chose to prioritise those aged between 15 and 24 and younger.

Frequencies and percentages of respondents indicating the age level that they thought the NHS should give priority to when compared with patients aged 65-74 who received the maximum amount of QALYs. (Rounds 1 and 2)

		Round 1 Frequency	Round 1 Percent	Round 2 Frequency	Round 2 percent
Valid	65-74 (QALY maximisation)	8	27.6	3	23.1
	Under 5	1	3.4	2	15.4
	5-14	1	3.4	0	0
	15-25	1	3.4	**3**	**23.1**
	25-34	0	0	1	7.7
	35-44	1	3.4	1	7.7
	45-54	1	3.4	1	7.7
	55-64	**16**	**55.2**	0	0
	Total	29	100.0	11	84.6
Missing	System	0	0	2	15.4
Total		29	100.0	13	100.0

Location of median respondent in bold.

Frequencies and percentages of respondents indicating the age level that they thought the NHS should give priority to when compared with patients aged 25-34 who received the maximum amount of QALYs. (Rounds 1 and 2)

		Round 1 Frequency	Round 1 Percent	Round 2 Frequency	Round 2 percent
Valid	25-34 (QALY maximisation)	9	31.0	5	38.5
	Under 5	3	10.3	**3**	**23.1**
	5-14	2	6.9	2	15.4
	15-24	**15**	**51.7**	1	7.7
	Total	29	100.0	11	84.6
issing	System	0	0	2	15.4
Total		29	100.0	13	100.0

Location of median respondent in bold.

Frequencies and percentages of respondents indicating the age level that they thought the NHS should give priority to when compared with patients aged 75-84 who received the maximum amount of QALYs. (Rounds 1 and 2)

		Round 1 Frequency	Round 1 Percent	Round 2 Frequency	Round 2 percent
Valid	**75-84** (QALY maximisation)	7	24.1	**7**	**53.8**
	65-74	**20**	**69.0**	4	30.8
	Total	27	93.1	11	84.6
Missing	System	2	6.9	4	15.4
Total		29	100.0	15	100.0

Location of median respondent in bold.

2.2.2.6 Children and dependents

When faced with a choice between a group of single adults gaining 1000 QALYs and a group of single parents gaining 900 QALYs with a varying number of children, the median respondent in Round 1, alongside the majority of respondents, would give priority to single parents irrespective of the number of children they have. QALY maximisation was clearly a minority view. On the other hand, in Round 2, a third of respondents support the QALY maximisation principle. The median respondent chose to give priority to single parents with two children.

Frequencies for the preference of each respondent for what they thought should be the minimum level that the NHS should prioritise when compared to patients that are single with no children but receiving maximum amount of QALYs. (Rounds 1 and 2)

		Round 1 Frequency	Round 1 Percent	Round 2 Frequency	Round 2 percent
Valid	Single 0 children (QALY maximisation)	2	6.9	4	30.8
	Single 3 children	1	3.4	1	7.7
	Single 2 children	0	0	**1**	**7.7**
	Single 1 child	**25**	**86.2**	5	38.5
	Total	28	96.6	11	84.6
Missing	System	1	3.4	4	15.4
Total		29	100.0	15	100.0

Location of median respondent in bold.

When the same question was asked in the context of dependents other than small children, similar results were obtained in Round 1. In Round 2, the median respondent chose to prioritise single patient groups with three dependents rather than single parents with two children as in the question above.

There was also a question that asked about single parents and couples with children. The first pairwise choice compared patients who were in a couple and had a child, and who will gain 1000 QALYs from additional treatment, with patients who were single and had a child, and who will gain 900 QALYs (or, 500 QALYs) from additional treatment. At the first choice, just under half the respondents in Round 1 chose the former group (QALY maximising), which is represented in the first three rows of the table, and the remaining chose to give priority to single parents. The QALY maximising respondents where then given pairwise choices where Group A was always the same as in the first pair while the number of children with the single parents (Group B) changed, and just under 30% of the whole sample remained QALY maximising regardless of the number of children the single parent had (1, 2, or 3). The other respondents who did not maximise QALYs at the first junction (represented in the following three rows) were given pairwise choices where those in Group B was always the same as in the first pair while the number of children for the couple changed. The modal preference for these respondents (and for the whole sample) was to give priority to the single parent with a child regardless of the number of children the couple had (1, 2, or 3).

Frequencies for the preference of each respondent for what they thought should be the minimum level that the NHS should prioritise when compared to patients that are single with no dependents but receiving maximum amount of QALYs. (Rounds 1 and 2)

		Round 1 Frequency	Round 1 Percent	Round 2 Frequency	Round 2 percent
Valid	Single 0 dependents, (QALY maximisation)	3	10.3	4	30.8
	Single 3 dependents	2	6.9	**4**	**30.8**
	Single 2 dependents	0	0	0	0
	Single 1 dependents	**23**	**79.3**	3	23.1
	Total	28	96.6	11	84.6
Missing	System	1	3.4	4	15.4
Total		29	100.0	15	100.0

Location of median respondent in bold.

Respondents in Round 2 were also required to prioritise those who were in a couple benefiting 1000 QALYs from the additional treatment and those that were single with one child that would benefit 500 QALYs from the additional treatment. For this question, over half (the first three rows) chose to maximise QALYs, whereas the remaining respondents choose to give priority to single parents. Those that choose to maximise QALYs were then presented with further pairwise choices, where Group A always consisted of patients in a couple that always benefited 1000 QALYs but Group B (benefited 500 QALYs) always consisted of single parents where the number of children changed (one, two or three children). Here, 50% of the full sample remained with the QALY maximising principle despite the number of children, whereas 10% chose to trade-off 500 QALYs when faced with the choice of prioritising single parents with two or more children and parents in a couple with one child. Those respondents that chose to prioritise Group B were presented with pairwise choices where Group B always consisted of single parents with one child and Group A comprised of those in a couple but the number of children changed (one, two or three children). The majority of these respondents (20%) chose to prioritise parents in a couple with two children over single parents with one child.

Frequencies for the preference of each respondent for what they thought should be the minimum level that the NHS should prioritise when compared to patients that are in a couple relationships with one child but receiving maximum amount of QALYs. (Rounds 1 and 2)

		Round 1 Frequency	Round 1 Percent	Round 2 Frequency	Round 2 percent
Valid	**Couple 1 child** (QALY maximisation)	8	27.6	**5**	**38.5**
	Single 3 children	1	3.4	0	0
	Single 2 children	4	13.8	1	7.7
	Prefer couple 2 children compared to single 1 child	1	3.4	2	15.4
	Prefer single 1 child to couple 2 children & couple 1child but not couple 3 children	2	6.9	1	7.7
	Prefer single 1 child compared to couple 1, 2 & 3 children	**12**	**41.4**	1	7.7
	Total	28	96.6	10	76.9
Missing	System	1	3.4	3	23.1
Total		29	100.0	13	100.0

2.2.3 Conclusion

A significant proportion of respondents supported QALY maximisation in the first set of focus groups where the difference between groups A and B was 100 QALYs. However, the median preference in most cases was to give the additional treatment to the disadvantaged group. The location of the median respondent was the same as the modal preference, except for the two social class questions. Except for the social class questions, the median and the modal preferences were to give priority to the disadvantaged at the smallest level of difference.

Building on our earlier finding that affordability matters, if the groups are defined in terms of income quartiles, 27% of respondents are in line with QALY maximisation (i.e. prioritise the “rich” who will get 1000 QALYs) whereas when the groups are defined in terms of social class 48% are consistent with QALY maximisation (i.e. prioritise social class 1 who will get 1000 QALYs). This seems to indicate that respondents may be looking at a concept of social class that is not solely determined by income.

The results from the focus groups that were presented with differences of 500 QALYs suggested that fewer respondents are willing to trade-off QALY maximisation, which can be attributed to the larger QALY difference between the groups. In the second set of groups, the most disadvantaged patient group was preferred by the median respondent in only one question (waiting lists) whereas, in the first set of groups, the median respondent was willing to sacrifice 100 QALYs across 1000 people to benefit the most disadvantaged patient groups in ten out of the twelve questions. These results suggest that respondents do broadly think about trade-offs between efficiency and equity.

2.3 STUDY C: RELEVANT ATTRIBUTES FROM POLICY-MAKERS

Study C is a survey of NHS staff, which was designed to inform the selection of the combinations of attributes by deriving weights for single and multiple attributes. Whilst this project is committed to the elicitation of the preferences of the general public, it is useful to inform the debate about which attributes should be valued by eliciting the preferences of those responsible for making resource allocation decisions and who will have to deal with the practical implementation of public preferences.

2.3.1 Methods

The questionnaire can be found in Appendix 2 and the attributes and levels used are summarised below. Respondents were asked about their personal preferences over who should receive health benefits on the basis of patient attributes (age, social class, length of time with illness, the presence of dependents, quality of life, and whether the illness was the result of NHS negligence); how relevant they believed the same patient characteristics to be in the allocation of scarce resources; and whether there were any reasons why policies based around these attributes would not be applied in the context of NICE and local decision making. The personal preferences, whilst potentially interesting in their own right, were elicited largely so that respondents could clearly distinguish these from the more important views (for our purposes) about the importance of these characteristics in a policy context.

The questions (attributes and levels) presented to respondents

Question on …	Programme A	Programme B
Age	5-25 years	60-80 years
Dependents	Without dependents	With dependents
Length of time with condition	Had for 1 year	Recently acquired condition
QOL without treatment	70%	40%
Cause of condition	Caused by NHS negligence	Caused by natural causes
Social class	Social class 1	Social class 5

For each characteristic, the response categories offered for the personal preferences questions were:

- I’d definitely prioritise programme A no matter what
- I’d probably prioritise programme A depending on what else is known
- I have no preference between the programmes
- I’d probably prioritise programme B depending on what else is known
- I’d definitely prioritise programme B no matter what

And for the corresponding resource allocation questions were:

- Would be relevant no matter what
- Would be relevant depending on what else is known
- Would not be relevant at all
- Policy makers would not want to know this info

The distribution of responses across the response categories was compared across attributes within the personal preference questions and within the policy relevance questions; and the distributions for the same attributes will be compared across the personal preference questions and the policy relevance questions.

A total possible sample of 1456 respondents was identified. At the macro level, senior officials in the Department of Health, NICE directors, senior persons from the Welsh Assembly, directors of the National Service Frameworks, and senior Directors of Public Health were all sampled (n=68). At an intermediate (meso) level, a database was compiled through the NHS web pages listing the members of the boards of the Strategic Health Authorities (SHAs) and Primary Care Trusts (PCTs). All 28 Chief Executives in SHAs, 298 chief executives in PCTs in England and 23 individuals who had titles such as head of resources or director of commissioning within PCTs were sampled. The micro level was represented by all 1039 clinicians active in the UK in the specialties of cardiac disease, cardio-vascular surgery, vascular surgery and respiratory medicine whose contact details were provided by a specialist web company.

A total of 380 questionnaires were returned after one reminder, which represents a 26% response rate. The majority of respondents were male (70%) and aged between 40 to 54 years (71%). Hospital clinicians made up nearly two-thirds of the sample, and most were from England. Two respondents had deleted their identification code, thus the remaining 348 were sent the postal questionnaire for the current survey. With an offer of entry to a prize draw, and one reminder, 172 completed replies were returned resulting in a 49% response rate (plus 9 uncompleted questionnaires).

2.3.2 Results

The table below shows that the 172 respondents were mostly male (66%), aged 40 to 54 (73%), with clinical training and currently holding a clinical post (71%). The majority of responses were in the middle ranges: indifference or a leaning towards one or the other programme in the personal preference question, and the "would be relevant depending on what else is known" category for the relevance for NHS decision making. The 'extreme' options (i.e. definitive preference for one programme over the other, or relevance or irrelevance to NHS decision making no matter what) were seldom the modal response.

Characteristic	Values	N	%
Sex	Males	113	66
	Females	39	23
Age	<39	8	5
	40-54	126	73
	55+	35	20
Training and post	Clinically trained, currently in clinical post	122	71
	Clinically trained, currently in managerial post	26	15
	Not clinically trained, currently in non-clinical post	18	0
	None of the above	3	2
Type of job	Hospital clinician	121	70

	PCT	34	20
	SHA/DoH	10	6
	None of the above	4	2

In the age question, one programme prioritised those aged 60 to 80 and the other those aged 5 to 25. The modal personal views question was not to prefer one group over the other, whereas the modal relevance response was to consider it relevant depending on what else is known.

How much would you be personally inclined to take **age** into consideration when making a choice over prioritising PA (5-25) or PB (60-80)?	How relevant do you think **age** would be in the decision to prioritise PA (60-80) or PB (5-25)?				
	Would be relevant no matter what	Would be relevant depending on what else is known	Would not be relevant at all	Policy makers would not want to know this info	Total
I'd def prioritise PA no matter what	1			1	2
I'd prob prioritise PA depending on what else is known	10	**50**	9	2	71
I have no preference between the programmes	4	47	17	6	**74**
I'd prob prioritise PB depending on what else is known	4	12	1	1	18
I'd def prioritise PB no matter what	1	1			2
Total	20	**110**	27	10	167

In the dependents question, one programme prioritised those with dependents and the other prioritised those without dependents. The modal personal response was no preference across the two programmes. The modal decision making view was that could be a relevant concern contingent on what else is known.

How much would you be personally inclined to take **dependents** into consideration when making a choice over prioritising PA (without dependents) or PB (with dependents)?	How relevant do you think **dependents** would be in the decision to prioritise PA (with dependents) or PB (without dependents)?				
	Would be relevant no matter what	Would be relevant depending on what else is known	Would not be relevant at all	Policy makers would not want to know this info	Total
I'd def prioritise PA no matter what		2			2
I'd prob prioritise PA depending on what else is known		12	6		18
I have no preference between the programmes	7	23	**35**	20	**85**
I'd prob prioritise PB depending on what else is known	6	34	10	12	62
Total	13	**71**	51	32	167

In the question on the length of time patients with a condition, one programme prioritised recent patients, whilst the other considered those had had the condition for one year. The modal personal response was that the respondent has no preference over the two programmes, whereas the modal relevance response was that the attribute would be of relevance depending on what else was known.

How much would you be personally inclined to take **length of time** with condition into consideration when making a choice over prioritising PA (had condition 1y) or PB (recently got condition)?	How relevant do you think **length of time** with condition would be in the decision to prioritise PA (recently got condition) or PB (had condition 1y)?				
	Would be relevant no matter what	Would be relevant depending on what else is known	Would not be relevant at all	Policy makers would not want to know this info	Total
I'd def prioritise PA no matter what	2	9			11
I'd prob prioritise PA depending on what else is known	8	**46**	8	4	66
I have no preference between the programmes	5	30	29	9	**73**
I'd prob prioritise PB depending on what else is known	2	6	5	1	14
I'd def prioritise PB no matter what		2			2
Total	17	**93**	42	14	166

For the question considering quality of life without treatment, one programme prioritised with 40% quality of life, and the other prioritised those with 70% quality of life. The modal personal preference is in favour of the latter group depending on what else is known. The modal relevance response was that untreated health would be relevant depending on what else is known.

How much would you be personally inclined to take **QOL** without treatment into consideration when making a choice over prioritising PA (70% QOL) or PB (40% QOL)?	How relevant do you think **QOL** without treatment would be in the decision to prioritise PA (70% QOL) or PB (40% QOL)?				
	Would be relevant no matter what	Would be relevant depending on what else is known	Would not be relevant at all	Policy makers would not want to know this info	Total
I'd def prioritise PA no matter what	3				3
I'd prob prioritise PA depending on what else is known	1	16	2	1	20
I have no preference between the programmes	2	30	14	5	51
I'd prob prioritise PB depending on what else is known	22	**52**	7	2	**83**
I'd def prioritise PB no matter what	5	2	2		9
Total	33	**100**	25	8	166

For the question on NHS negligence, the modal personal preference was neutral, whereas the information would be relevant to NHS decision-making depending on what else was known.

How much would you be personally inclined to take **NHS negligence** into consideration when making a choice over prioritising PA (NHS negligence) or PB (naturally caused)?	How relevant do you think **NHS negligence** would be in the decision to prioritise PA (NHS negligence) or PB (naturally caused)?				
	Would be relevant no matter what	Would be relevant depending on what else is known	Would not be relevant at all	Policy makers would not want to know this info	Total
I'd def prioritise PA no matter what	9	1	2	1	13
I'd prob prioritise PA depending on what else is known	20	28	6	5	59
I have no preference between the programmes	9	26	**31**	20	**86**
I'd prob prioritise PB depending on what else is known		2	4	1	7
I'd def prioritise PB no matter what		1		1	2
Total	38	**58**	43	28	167

Regarding the question on social class, where one programme prioritised those in social class 1 and the other programme prioritised those in social class 5, the (overwhelming) modal personal preference was indifference between the two programmes, whereas the modal preference for NHS decision making was that it would be relevant depending on what else is known.

How much would you be personally inclined to take **social class** into consideration when making a choice over prioritising PA (social class1) or PB (social class 5)?	How relevant do you think **social class** would be in the decision to prioritise PA (social class1) or PB (social class 5)?				
	Would be relevant no matter what	Would be relevant depending on what else is known	Would not be relevant at all	Policy makers would not want to know this info	Total
I'd def prioritise PA no matter what		1			1
I'd prob prioritise PA depending on what else is known	1	1	2		4
I have no preference between the programmes	8	26	**49**	36	**119**
I'd prob prioritise PB depending on what else is known	6	25	3	2	36
I'd def prioritise PB no matter what	3	2		1	6
Total	18	**55**	54	39	166

2.3.3 Conclusions

The results show that when the personal preference questions are asked, the predominant modal response is not to have a preference to favour one patient group over the other, when patient groups differ in terms of their age, dependents, the length of time they have had the condition, NHS negligence, and social class. This is despite there being an option to prioritise one programme over the other ‘depending on what else is known about the patients’. Quality of life without treatment was the only attribute where this category was the modal response. It is interesting to note that some of the patterns found here are in contrast to what we have found in studies A and B from the members of the public. For instance, there was no strong preference to support prioritising the young, lower social class or victims of NHS negligence.

For the ‘policy relevance’ questions, the modal response across all attributes is to say it would be relevant depending on what else is known. As a result, for NHS decision making, attributes are in general considered relevant depending on other attributes, which provides some support for studying multi-attribute weights, as opposed to single attribute weights.

The data show that not everybody has avoided the somewhat extreme ‘no matter what’ option. Furthermore, data show that the proportion of those taking this extreme option varies across questions, thus indicating that the question captures varying strengths of views across the attributes and perspective.

2.4 STUDY D: DISPERSION VS. CONCENTRATION OF BENEFITS

If two health programmes both produced 100 person years, can we say their benefits are the same? Conventional economic evaluation of health care interventions assumes that they are equivalent. On the other hand, there seems to be some intuition based on inequality aversion that people may wish to disperse health benefits rather than to concentrate them to a few, so that for instance an additional 5 years each to 20 people seems more preferable than an additional 50 years to just 2 individuals. However, it also seems absurd that for instance extending the life of a very large number of people, say about half a million, by one minute each should be regarded as equivalent to extending the life of one person by one year (365 days × 24 hours × 60 minutes = 525,600). In other words, dispersing health benefits beyond some limit, or a threshold, will probably diminish the value of the total benefit.

Patterns of dispersion and concentration have been found in empirical studies by Ubel et al (1996), Choudhry et al (1997), and Ubel et al (2000). The issue of specific utility functions and the level of thresholds were first explicitly explored in the heath field by Olsen (2000; also see Olsen, 1994), and subsequently Rodríguez-Míguez, Pinto-Prades (2002) proposed a method to actually identify the threshold level. Preparatory study 3 is a quasi-replication of the study by Rodríguez-Míguez and Pinto-Prades (hereafter referred to as the RP study). In what follows, the RP study is briefly introduced, the changes made in the current study are explained, and then the results are reported.

2.4.1 Methods

2.4.1.1 Questionnaire design

The original RP study assumes that social welfare is a sum of individual utility from QALY gains, and where individual utility from marginal QALY gains are not constant. More specifically, the individual utility function is assumed to take the form

$$u(t_i) = \alpha_1 \exp(-\alpha_2 t_i) t_i^{\alpha_3}, \qquad [1]$$

where u is utility and t_i (i = 1,2, … n) represents QALY gains to individual i, so that both positive and negative inequality aversion can be expressed. Positive inequality aversion is a preference for the dispersion of a fixed total benefit, and is associated with a negative value of $u''(t)$. Correspondingly, negative inequality aversion is a preference for the concentration, associated with a positive $u''(t)$. Thus, by exploring the value of t above which $u''(t)$ is negative and below which it is positive, the threshold value can be identified.

Imagine a reference programme that gives 10 additional years of life in full health to 10 people, denoted (10 years, 10). If inequality neutrality holds, then with zero discounting this programme and the following 100-person-year scenarios should produce the same level of individual utility and thus social welfare: (1 year, 100), (2 years, 50), (5 years, 20), (20 years, 5), (50 years, 2). However, if there is positive (negative) inequality aversion, then for the reference programme to be equivalent to each of these scenarios the number of people (p^*) each benefiting by 10 years has to be larger (smaller) than 10. This is in effect a person trade-off exercise, and the RP study uses “choice bracketing” to operationalise this. The objective of the exercise is to identify the value of p^* that will make a respondent indifferent between a given scenario and (10 years, p^*). If inequality neutrality holds, then (with zero discounting) $p^* = 10$ should hold.

In order to operationalise the person trade-off exercise and to identify the value of p^* for each scenario, the following series of 10-year reference programmes is set: (10 years, 1), (10 years, 3), (10 years, 5), (10 years, 8), (10 years, 10), (10 years, 12), (10 years, 15), (10 years, 18), (10 years, 20). These reference programmes do not generate the same sum of health benefits, and, provided the respondent supported the view that social welfare is increasing in the number of people treated, there is a dominant ordering so that the programmes further down the list are more preferable. By bracketing and narrowing down the number of people in the reference programme, this will lead to the identification of the value of p^*. Furthermore, if the value of p^* varies across the 100-person-year scenarios, then it will be possible to infer for each respondent a ranking of the scenarios by using the size of p^*.

An important assumption at this point is that $u(t) \times p = u(10) \times p^*$, where t and p each represent the number of years and persons in the scenario in question. By solving this for $u(t)$,

$$u(t) = 10 \times p^* / p, \qquad [2]$$

which allows the use of regression analysis to estimate the coefficients of equation [1], and in turn the identification of the threshold value.

We made seven main changes to the design used in the RP study. Firstly, whereas the respondents of the RP study are students, the respondents of the current study are members of the general public. Secondly, the RP study contacted the respondents over three sessions: session one to ensure respondents understood the exercise, session two to carry out the main exercise, and session three to test for reliability over time. However, respondents in the current study each attended one session only.

These two points lead us to think the “choice bracketing” method used in the RP study to identify indifference between scenarios may be too confusing. This consists of a series of choices between pairs of treatments laid out in table format, alongside instructions on what the respondent should do next depending on the answer for the choice in question (see Appendix of the RP study for a replication). Therefore, thirdly, a new mode of administration, based on a set of cards with different treatment pairs was used. Each 100-person-year *scenario* was treated as one *question*.

Within each question, a set of nine cards was used. (See Appendix 3 for an example.) Each card had two treatments printed on them; “treatment A” represented the *scenario* in question and thus remained the same across all cards within a question; “treatment B” represented one of the nine 10-year *reference programmes*, taken from the ordered list explained above. For each question, respondents were asked to look at the set of cards and place them in three different batches: one where treatment A is better than treatment B, one where two treatments are equally good, and one where treatment B is better than treatment A. An advantage of this mode of administration is that there is ample opportunity to test for a respondent’s consistency within and across questions.

The fourth change concerns the scenarios. The RP study used five 100-person-year scenarios presented above, with the number of years ranging from 1 to 50. However, it was felt that shorter durations needed exploring as well, and thus two scenarios (9 months, 150) and (6 months, 200) was added, and the scenario (2 years, 50) was dropped. See the next section for further details about the questions asked at different stages of the study. Fifthly, the RP study assumed that all patients were 20 year-olds. However, in order to explore whether the preference for dispersion and concentration is independent of patient age, in addition to the six main questions assuming all patients are 20, two further questions were introduced, one where patient age was set at 10, and another at 60. The 100-person-year scenario (5 years, 20) was used in both questions.

The person trade-off is a method that uses the number of people as the vehicle of trade-off, and it may be affected by the level of this numéraire. More specifically, when a respondent prefers a treatment involving “5 patients” does this figure have an absolute value, regardless of whether it is 5 patients out of a pool of 200, or a pool of 2000; in other words would the reference group have an effect? Furthermore, would the same patient who chose 5 patients out of 200 also choose 50 patients out of 2000; in other words, would constant proportional *person* trade-off hold? The sixth change explored these issues. All questions explicitly stated that all those treated were drawn from a pool of 200 patients (and the number of those treated in the scenarios ranged from 2 to 200), except for two additional questions. One of them stated that the pool of patients was 2000, but all other figures remained the same: this will test for the reference group effect. The other scaled up all the numbers of people (i.e. the pool of patients and the number of those treated; and thus the person-years) by 10: this will test for constant proportional person trade-off.

One concern over the RP design was that it used years of life in full health, and not QALYs. 100 person-years in full health equals 100 QALYs. And not only can 100

QALYs be made up of different combinations of years of life in full health (e.g. 5 years to 20 people; 10 years to 10 people; etc), if the QALY concept holds in this context, each of them should be equivalent to different combinations of number of years, health related quality of life, and numbers of people (e.g. 5 years in full health to 20 people; 5 years in 50% QOL to 40 people; 10 years in 50% QOL to 20 people; etc). However, it was considered to be too complicated to explore this issue in detail. Instead, the seventh change consisted of a simple test of whether or not the preference for dispersion or concentration was affected when the outcome of the two treatments were less than full health (but the same across the two treatments). One question was added, where the scenario in question was specified as (40 years, 50% QOL, 5), which is comparable to (20 years, 5). Furthermore, three questions were introduced which are: (20 years, "slight mobility problems", 5), (20 years, "slight pain", 5), and (20 years, "mild depression", 5).

2.4.1.3 Recruitment and respondents

This is a quantitative study and we sought to elicit responses from about 60 members of the general public. This was not based on any power calculations as such but, given the question type, is consistent with the sort of samples for which statistically significant results can be generated. The respondents were recruited by approaching people in Sheffield. The survey was carried out in group sessions of about 10 respondents each held in University seminar rooms. The sessions began with some general descriptive information about the nature of scarce health care resources and the need for prioritisation at a macro level; and how current policy is based on concerns for maximising person years with no particular concern for the distribution of it across identical patients.

Respondents were then broken up into smaller groups of 2 to 5, each lead by a member of the research team. This facilitated closer monitoring of how the respondents progressed with the exercise. The respondents were given a questionnaire booklet consisting of 16 questions across 16 pages, accompanied by corresponding sets of 9 cards as presented above, and an explanation of how the questionnaire worked. Table 35 summarises the questionnaire. To avoid confusion, the cards and the pages of the booklet were colour-coded so that for instance all the green cards were used for the question printed on the green page, and so on. Each page in the booklet had 3 boxes printed: 'I prefer treatment A', 'Treatments A and B are equally good' and 'I prefer treatment B', so that respondents can use these to group the nine cards as explained above. All cards were placed in a random order and respondents were asked to shuffle the cards each time they were used. Once respondents completed this for all 9 cards they were then advised that they could swap the cards into different boxes as they saw fit and then asked to write down which cards they had place in each box onto the booklet.

Once all respondents completed this for the first question, the whole group moved to the next question, and the same steps were repeated for the following 11 questions. After question 11 the group had a brief open discussion about the questionnaire and the topic. Once all smaller groups reached this point, a brief explanation was given to the wider group on the notion of health related quality of life, and how living for a longer number of years and poorer health might be equivalent to living for a shorter duration in better health. Respondents then broke up into the smaller groups and followed the above steps for remaining questions 12, 13, 14 and 15.

After these person trade-off exercises, question 16 was a direct ranking exercise, where respondents were provided with 7 cards each representing the 100-person-year scenarios used in the 6 person trade-off exercises, and the reference programme (10 years, 10). Respondents were asked to rank them from 1 (most preferred) to 7 (least preferred). The aim of this exercise was to analyse the extent to which the implied ranking obtained from the person trade-off will coincide with the results of the direct ranking.

Questions asked to respondents and different variants used.

Different variants	Treatment A (Health gain, patients)
Baseline variant:	
Question 1	(6 months (0.5years),200)
Question 2	(9 months (0.75years),200)
Question 3	(1 year, 100)
Question 4	(2 years, 50)
Question 5	(5 years, 20)
Question 6	(20 years,:5)
Question 7	(50 year, 2)
Age variant:	
Question 8. All patients aged 60	(5 years, 20)
Question 9. All patients aged 10	(5 years, 20)
Pool size variant:	
Question 10. Pool size = 200	(5 years, 200)
Question 11. Pool size = 2000	(5 years,200)
Less than full health variant:	
Question 12. Health gain means living in a 0.5 QOL	(20 years,5)
Question 13. Health gain means living with slight mobility problems	(20 years, 5)
Question 14. Health gain means living with slight pain	(20 years, 5)
Question 15. Health gain means living with mild depression	(20 years, 5)

Since, to the best of our knowledge, this was the first time this card method had been used to measure the indifference value of concentrating or dispersing benefits, the design was adapted throughout the sessions. As the length of each session was unknown during the first 2 sessions (N=20) respondents were given only 9 questions: (1 year, 100), (2 years, 50), (5 years, 20), (20 years, 5), (50 years, 1), (5 years, 20), the age 60 question, the age 10 question, (5 yeas, 200) out of a pool of 200, (5 yeas, 200) out of a pool of 2000. In other words respondents answered questions 3 – 8.

In the third session (N=12) 5 questions were added: (0.5 years, 200), (40 years, 50% QOL, 5) and 3 questions on scenarios with specific health problems; and instead (2 years, 50) (i.e. question 4) was removed. A further direct rank ordering exercise (question 16) was added. The 4 cards ranked included: (5 years, 20), (20 years, 5), (50 years, 2), (10 years, 10). In the final 4 sessions (N=26) an extra question on (0.75 years, 150) was also included to gage more fully people's preferences over smaller *t*'s. Question 16 expanded to include all the scenarios.

There were opportunities throughout the meetings for respondents to voice any opinions that they might have had regarding the task. Sessions lasted between 60 and 90 minutes, depending on the number of questions, the speed of the respondents and amount of queries and opinions voiced throughout the meeting. Sessions were audio-taped with the respondents consent and brief notes were taken throughout to aid clarity when interpreting the results. At the end, respondents were asked to fill in a background questionnaire, and to write down their thoughts regarding the content of the group sessions. Respondents were given a payment of £15 for participating in the study and thanked.

2.4.1.3 Analysis

The analysis involves attaching a score for each 100-person-year scenario by respondent corresponding to the position relative to the ordered list of 10-year reference programmes. Observations have to be dropped if these cannot be assigned unambiguously due to inconsistency. There are three kinds of inconsistencies. First, "*across* question inconsistency" arises, if a respondent places all cards in the box where treatment A (or B) is preferred throughout the main questions: it is not inconsistent to say that for example scenario (2 years, 50) is better than any of the reference programmes (e.g. because only the number of people matter, and not person years), but then this will be inconsistent with saying that scenario (50 years, 2) is also better than any of the reference programmes. In such cases all observations from the same respondent is removed.

Second, "*within* question inconsistency" arises, if a response indicates that a given scenario was better than say (10 years, 12) but worse than (10 years, 8), and this observation has to be excluded. However, when the same apparently inconsistent response was given in the less than full heath questions, these were not excluded, since they are in line with the maximal endurable time hypothesis (Sutherland et al, 1982; Stalmeier et al, 1996), and not necessarily illogical. Third, there were several questions where another type of within question consistency could be tested, if respondents chose treatments that had less health benefits rather than treatments that had more health benefits to the same number of people; for example Treatment A is preferred when Treatment A is (5 years, 20) and Treatment B is (10 years, 20). Individual responses with this type of inconsistency will be excluded.

For each individual respondent, an implied ranking of the 100-person-year scenarios were obtained from the results of the main person trade-off exercises, and this was compared against their own direct ranking of the scenarios using Spearman rank order correlations. The mean coefficient across respondents is reported. Note that some respondents (*N*=12) give ranking data for 5 scenarios and others (*N*=26) for 7. For each scenario, the average implied rank score and the average direct rank score were calculated across individual respondents, and the Spearman and the Kendal correlation coefficients are reported.

The value of p^* was identified for each observation. When respondents indicated that treatment A was better with respect to some 10-year reference programmes and treatment B was better with respect to other references, but did not directly reveal an indifference value, or if they demonstrated that they were indifferent on more than one programme, the intermediate value of the interval was used. If respondents always chose pairs with the greatest number of patients or the pairs with the greater number of years (i.e. thought treatment A or alternatively treatment B was better in all instances) an indifference value was inferred (i.e. in questions 1-9 and 12-15 if treatment A was always preferred $p^* = 20$ and if treatment B was always preferred $p^* = 1$; in questions 10 and 11 if treatment A was always preferred $p^* = 200$ and if treatment B was always preferred $p^* = 10$). This inferred indifference value was used when respondents did not make trade-offs. As was explained above, the crucial value of p^* with zero discounting is 10, and thus a series of 2-sided t-tests are carried out for each question. If mean $p^* = 10$, then the implication is that inequality neutrality holds and that on average people are indifferent between the scenario in question and the reference programme (10 years, 10).

Each scenario considers the concentration or dispersal of 100-person years, and the average preference over these options is reported. For a given pair of scenarios, if the scenario with the larger number of persons is associated with a larger mean p^*, then this is regarded as a preference for dispersion. See for instance the pair (1 year, 100), where mean p^* is 15.49, and (5 years, 20), where mean p^* is 12.85. The former scenario has a larger number of persons (100 > 20) and a larger mean p^* (15.49 > 12.85), and therefore the cell representing this pair of scenario is marked with a “D“” dispersion. The opposite relationship holds for concentrating preferences. Finally, values of $u(t)$ at the individual level are calculated using equation [2] and based on individual values of p^*. This is used to estimate equation [1], and subsequently to identify the level of t where $u''(t) = 0$, which is the dispersion-concentration threshold.

The effect of patient age, the effect of reference groups and constant proportional person trade-off, and the effect of less than full health are explored in terms of mean p^* and t-tests. Note that while these can establish whether or not the preference between concentration and dispersion observed at the main questions also apply at these different settings, we cannot draw conclusions about the robustness of the dispersion-concentration threshold value from these results.

2.4.2 Results

The study was conducted on members of the general public recruited in Sheffield. Out of the 96 individuals that agreed to attend the sessions 68 respondents (71%) actually took part across seven sessions. The background details of the respondents for whom background data are available are shown in the table below

The background questionnaire also listed five negative and five positive adjectives used in studies A and B that may describe the experiences and thoughts that the respondents might have of the session, and invited respondents to choose as many as they thought applied. Results showed that the majority of respondents thought that the sessions were: “interesting” ($N = 57$); “insightful” ($N = 32$); “informative” ($N = 20$); “enjoyable” ($N = 21$); and “too sensitive a topic” ($N = 7$).

Of the 68 respondents, 49% had no inconsistencies, and a further 24% only one inconsistency throughout the exercise. However, two respondents (2.94%) were inconsistent across questions and were excluded from the analysis, and a further eight respondents (11.8%) violated the first type of within question inconsistency on more than eight questions, and were also dropped. In terms of specific observations, six responses to questions (5 years, 20), (5 years, 20, age 10) and (5 years, 200, out of 2000) were excluded due to the second type of within question inconsistency.

Background questions	Number of respondents (total n = 67)
Sex: Male	28
Female	39
	(Mean = 35.3, SD = 0.50)
Employment Status:	
Full-time work	10
Part-time work	14
Student	30
Seeking work	2
Home-maker	2
Retired	9
Highest qualification:	
GCSE/O level	6
NVQ	5
A level	20
Degree	15
Postgraduate qualification	11
No formal qualifications	9
Has previously or presently worked for NHS	12
Has previously or presently been a main carer for a disabled family member	12
Children:	
0 children	51
1 child	5
2 children	4
3 children	1
Perceptions of own QOL ranged between 40% and 100% (Mean = 85.42, Median = 90.00, SD = 13.89)	

The next table summarises the mean value of *p** derived from the main questions for the 100-person-year scenarios, and the *t*-test results. They suggest that the inequality neutrality assumption is rejected, so that on average people are concerned about how a fixed benefit of 100 person years is distributed across the pool of 200 patients. The results also illustrate that the value of mean *p** starts below ten persons for (6 months, 200) increases to around 15 persons at (1 year, 100), and then declines to around five persons, with zero discount. The pattern for durations over one year is consistent with positive time preference, but not the pattern under one year.

The table following this summarises combinations of scenarios, and indicates whether respondents preferred dispersed or concentrated benefits. For example, increased dispersion is preferred in a choice between splitting 100 person years amongst 20 people (five years each) or 50 people (two years each).

Assessment of health programmes

Health gain (years), *t*	Number of patients, *p*	*Mean p**[a] (2-sided *t-test*)[b]
0.5	200	9.07 (0.25)
0.75	150	11.32 (0.310)
1	100	15.49 (0.001)
2	50	14.83 (0.001)
5	20	12.85 (0.001)
10	10	10
20	5	8.32 (0.001)
50	2	5.52 (0.001)

[a] Number of patients who would have to receive a 10 life-year increase in order that this programme be indifferent to the (*t,p*) programme.

[b] H_0: mean $p^* = 10$; mean $p^* \neq 10$; $n = 58$.

Distributive preferences based on individual gain (years) [a]

(Gain in years, patients)	0.75, 150	1, 100	2, 50	5, 20	10, 10	20, 5	50, 2
0.5, 200	*C*	*C****	*C****	*C****	*C*	*D*	*D****
0.75, 150		*C****	*C***	*C*	*C*	*D****	*D****
1, 100			*D*	*D****	*D****	*D****	*D****
2, 50				*D**	*D****	*D****	*D****
5, 20					*D****	*D****	*D****
10, 10						*D****	*D****
20, 5							*D****

[a] *C:* preferences for concentrating; *D:* preferences for dispersing

***significance at 1% level; ** significance at 5% level; *significance at 10% level; *n* = 58.

Equation [1] is estimated as a linear function, resulting in:

$$\hat{u}(t) = 0.941\exp(-0.029t)t^{1.173}.$$

The value of *t* at which the second derivative of this equation is zero represents the threshold value between dispersion and concentration of benefits, and corresponds to *t* = 2.6 years (assuming no temporal discounting; with positive discounting this value will be higher). This value is considerably smaller than the comparable value reported in the RP study, which is 9.1 years.

The table below shows that the values of *p** are statistically significantly larger than 10 across the three ages: in other words, if the choice is between giving 5 years to a larger number (20) or giving 10 years to a smaller number (10) then the preference to disperse rather than to concentrate is robust. There also seems to be a pattern across the three ages so that this preference is stronger when patients are older, but the results for 10-year olds and 60-year olds are not statistically significantly different from the results for 20-year olds.

The effect of patient age

Gain in years, *t* (age of patients)	Number of patients, *p*	*Mean p**[a] (2-sided *t-test*)[b]	Effect of age (2 sided *t*-test)[d]
5 (age 60)	20	13.84 (0.001)	(0.49)

5 (age 20)[c]	20	12.85 (0.001)	-
5 (age 10)	20	12.55 (0.001)	(0.82)

[a] Number of patients who would have to receive a 10 life-year increase in order that this programme be indifferent to the (t, p) programme ($n = 58$).
[b] H_0: mean $p^* = 10$; mean $p^* /=10$; $n = 26$.
[c] This row represents the corresponding row in Table 43 above.
[d] H_0: mean p^* = mean p^* for age 20

As can be seen in the second and third rows of the following table, values of p^* are statistically significantly larger than 100 across the two additional questions: in other words, if the choice is between giving 5 years to a larger number (200) or giving 10 years to a smaller number (100) then the preference to disperse rather than to concentrate is robust across different patient numbers overall. The comparison between the second and third rows represents the effect of reference group: i.e. does it matter whether the 200 patients are from a pool of 200 or from a pool of 2000? The comparison between the first and the third rows represents constant proportional person trade-off: i.e. if all relevant numbers are scaled up by ten, would the value of p^* also be scaled up by ten? The results cannot reject that constant proportional person trade-off holds.

The effect of different frames

Health gain (years), t	Number of patients, p	*Mean* p^*[a] (2-sided *t-test*)	Effect of frame (2 sided t-test)
5[c]	20 out of 200	12.85 (0.001)[b]	-
5	200 out of 200	124.15 (0.034)[d]	(0.579)[e]
5	200 out of 2000	131.72 (0.007)[d]	(0.756)[f]

[a] Number of patients who would have to receive a 10 life-year increase in order that this programme be indifferent to the (t, p) programme.
[b] H_0: mean $p^* = 10$; mean $p^* /=10$; $n = 26$.
[c] This row repeats the corresponding row in Table 43 above.
[d] H_0: mean p^* = mean $p^* \times 10$ for 20 out of 200
[e] H_0: mean p^* = mean p^* for 200 out of 2000
[f] H_0: mean p^* = mean p^* for 20 out of 200

The table below illustrates that when the total health benefit of 20 years in full health to 5 persons is substituted with 40 years in 50% health to 5 persons, p^* remains statistically significantly smaller than 10: i.e. to disperse across 10 persons than to concentrate on 5. The last column indicates that the two values of p^* are not significantly different from each other, so we preferences for dispersion of health benefits in year of life in full health might be generalised to health benefits in QALYs, provided the composition of QALYs is equal across the two alternatives (i.e. in this case, all treatments in this question had 40 years in 50% QOL).

The effect of less than full health

Health gain (years), t	Number of patients, p	*Mean* p^*[a] (2-sided *t-test*)	Effect of QOL adjustment. (2 sided t-test)

20[c]	5	8.32 (0.001)	-
40 (1/2 QOL)	5	7.33 (0.001)	(0.10)[d]

[a] Number of patients who would have to receive a 10 life-year increase in order that this programme be indifferent to the (*t, p*) programme.

[b] H_0: mean $p^* = 10$; mean $p^* \neq 10$; $n = 26$.

[c] This row repeats the corresponding row in Table 43 above.

[d] H_0: mean $p^* = p^*$ for (20 years, full health, 5)

As can be seen in the three bottom rows of the following table, the values of p^* are statistically significantly smaller than 10 across slight mobility problems, slight pain, and mild depression; in other words, if the choice is between giving 20 years with a given health problem to a smaller number (5) or giving 10 years with the same health problem to a larger number (10) then the preference to disperse rather than to concentrate is robust across these different health problems. The last column indicates that the values of p^* from the three questions is not different from that obtained in the full health case.

The effect of different types of ill health

Health gain (years), *t*	Number of patients, *p*	*Mean* p^*[a] (2-sided *t-test*)	Different health problems. (2 sided *t*-test)
20[c]	5	8.32 (0.001)	-
20 (slight mobility problems)	5	8.47 (0.001)	(0.81)[d]
20 (slight pain)	5	7.69 (0.001)	(0.14)[d]
20 (mild depression)	5	7.84 (0.001)	(0.40)[d]

[a] Number of patients who would have to receive a 10 life-year increase in order that this programme be indifferent to the (*t, p*) programme.

[b] H_0: mean $p^* = 10$; mean $p^* \neq 10$; $n = 26$.

[c] This row repeats the corresponding row in Table 43 above.

[d] H_0: mean $p^* = p^*$ for (20 years, full health, 5)

2.4.3 Conclusions

Current decision-making criteria do not take into consideration how a given fixed total health benefit is distributed across a group of people. If several treatments all generate 100 QALYs, then their benefits are assumed to be equivalent. However, this practice may not mirror the preferences held by members of the public. It has been pointed out that (i) people may prefer to disperse the fixed benefit across a larger number of patients rather than to concentrate it to a smaller number of patients, but that (ii) this may be subject to a threshold so that the size of health gain to an individual patient is not too small.

Different values of p^* were observed for different 100-person-year scenarios, indicating that the assumption of distribution neutrality does not hold. On average, respondents preferred to disperse the 100 person years but not when the size (duration) per patient was small; this threshold was identified at 2.6 years (with zero discounting). The results of the additional questions were generally favourable towards the general assumptions of economic evaluation: none of the following changes resulted in statistically significantly different values of p^*: changing the age of patients from 20 years old to 10 years old or to 60 years old; changing the size of the reference group from 200 to 2000; scaling up the whole scenario by ten; and using various kinds of less than full health.

2.5 OVERALL CONCLUSIONS FROM PHASE ONE

This chapter has reported on the first phase of the project, which was broken down into four studies. Studies A and B sought to identify the relevant attributes and to identify meaningful levels of these attributes from focus groups with members of the general public. The attributes considered were age, social class, length of time with condition, dependents, quality of life without treatment, and whether the condition was caused by NHS negligence. All of these attributes were considered relevant in Study A, and respondents chose to use these characteristics on the basis of which to divert from simple QALY maximisation.

While there is a growing literature on the attributes across which people may have preferences to over-ride QALY maximisation, the advantage of Studies A and B were that multiple attributes were examined within the same study design. Some of the findings pointed to the complexity of the issues. For example, when people chose to give lower priority to those in higher social classes, one reason for this was because they will be able to purchase private care, outside the NHS. Similarly, the issue of dependents was found to be complex because it was not simply whether or not a patient had any dependents but also the age and number of the dependents, whether the patient was the only adult responsible for them, and so on.

All the attributes examined in Study A were addressed in Study B, which aimed to identify the minimum level where the attribute begins to become relevant. For example, people may feel social class in general to be a relevant characteristic in health care priority setting, and yet not distinguish between two social classes adjacent to each other. The exercise built on that used in Study A but there was less debate, and the respondents' preferences were elicited in a quantitative manner. In addition, some of the questions were broken down further to address the complex issues raised above. For example, an income group question was inserted before the social class question, and the social class question was asked in two contexts (i.e. with and without health care available outside the NHS). Further details around having dependents were also added.

Study B was carried out in two separate rounds, which varied the number of QALYs that would have to be sacrificed to target a priority group. As expected, when the opportunity cost of diverging from QALY maximisation was larger, a smaller proportion of respondents were ready to do so. Therefore, if people are asked about their preferences to prioritise certain patient characteristics without specifying what the associated opportunity costs are, they may be found to diverge from QALY maximisation more often than in situations where the opportunity costs are made more explicit.

Attributes for which respondents were willing to sacrifice health gains when the sacrifice was 500 QALYs as well as when it was 100 QALYs were: quality of life without treatment (when the lower level was 40%), time with condition (where one party had the condition for one year or more), time on waiting list (one month or more), age (15-24 year olds relative to 65-74 year olds; and under-fives relative to 25-34 year olds), and dependents (single adults with 2 or more children).

Study C used NHS staff as respondents, with the aim to identify the most policy relevant combinations of attributes. Questions were asked about personal preferences as well as about the policy relevance of the attributes. A large survey followed three qualitative telephone interviews. The survey results suggest that the NHS staff sample have less strong personal preferences to divert from QALY maximisation than members of the public (subject to the important caveat that neither sample was entirely representative of the population from which it was drawn).

The most common result across all attributes was that the attribute is relevant to NHS decision-making depending on what else is known. No single attribute was regarded by a majority as relevant to NHS policy decision making ‘no matter what’. The three attributes where the majority of respondents thought the attribute in question ‘would be relevant depending on what else is known’ are: age (5-25 as opposed to 60-80), length of time with condition (more than one year as opposed to recent) and quality of life without treatment (40% as opposed to 70%). The only attribute that attracted a fifth of respondents indicating it ‘would be relevant no matter what’ was NHS negligence.

Based on the results of studies A, B and C, three attributes were taken forward into later phases of the project: age, quality of life without treatment and responsibility. Age was clearly important but there was no real consensus about when age mattered for policy and when it did not – except in the case of children versus adults, which is therefore the focus of our comparisons. For quality of life without treatment, a value of 25% was chosen for comparison with full health as this we can be confident that this meets the threshold for the general public to treat quality of life as significantly lower than full health (many respondents did distinguish better 70% and 40% but this was not a unanimous view).

In terms of responsibility, NHS negligence appears to be important to both NHS staff and the general public, especially in the context of patient safety. If NHS responsibility is included, then it does not make sense to have NHS responsibility versus everything else, because ‘everything else’ is too wide-ranging: but, equally, it is not reasonable to say that something is entirely within an individual’s control. Therefore, it was decided to present this attribute as three categories: “not the result of their lifestyles but is instead caused by errors within the NHS” (NHS responsibility, no patient responsibility); “caused by a combination of factors including poverty, genetics, pollution, and the patients’ lifestyles” (no NHS responsibility, limited patient responsibility); “caused by a combination of factors including poverty, genetics and pollution, but is not caused by patients’ lifestyles or by NHS error” (no NHS responsibility, no patient responsibility).

In addition, we have also considered the rarity of the condition. This attribute was added after NICE had requested that we also include ultra-orphan drugs due to their increasing prominence in recent policy debates. The definition of ultra-orphan conditions is often a mixture of the severity of the condition, the non-existence of alternative treatments, the immediacy of possible death without treatment and the rarity of the condition (McCabe et al, 2005). The severity of the condition is covered in our study but the non-existence of alternative treatments is far too detailed a level of analysis for this kind of macro level project and the immediacy of death is difficult to capture in a lifetime health framework. Therefore, we focused on rarity. Note, however, that this was not considered at Phase 1 and we have no data as to the reasons individuals may give more (or less) weight to increasingly rare conditions.

Three attributes were dropped: dependents, length of time with the condition and social class. The main reason for dropping these attributes is the complexity associated with each of them. For instance, the results from our studies suggest that it is not the status of having dependents per se that carries weight but the individual situation of being the sole responsible adult for small children. This is a level of detail that it would be difficult to reflect in NICE-level decision making, unless it were the case that being the sole responsible adult for small children was a key risk factor for a particular disease, or a central patient characteristic for the indication for a specific treatment, which are both unlikely. Furthermore, according to our understanding of the NICE methods guide, the effect of ill-health on others can be submitted as part of the appraisal process. In other words, this is an efficiency issue rather than an equity issue. As such, it may not be appropriate to include this within our study (although it may feature indirectly in people’s preferences over age).

Similarly, the studies found that the length of time with condition could have different implications depending on whether it was the length of time since acquiring the condition, or the length of time since being put on a waiting list. The length of time since acquiring the condition is to some extent a matter of individual circumstances including patient preferences, and, as with the case of dependents, a level of detail not relevant to NICE decision making. The issue of length of wait on the waiting list can to some extent be re-interpreted under NHS responsibility. Furthermore, whilst length of time with condition was believed to be important by NICE and NHS, this type of factor is likely to affect relatively few individuals, and is most likely to affect only those cases where a new type of treatment is developed, where the “starting point” for the wait is ambiguous. Weights at the individual level (e.g. between GP referral and inpatient treatment) is not really an issue for NICE, which deals with particular treatments or patient groups, not individual patients.

Whilst social class was important to respondents, it is very difficult to distinguish between any weighting on the basis of social class per se, quality of life (which we consider) and the number of quality-adjusted life-years (QALYs) enjoyed throughout a lifetime (which is largely captured within the inequality aversion parameter), and the impact of income and the availability of private health care. We are aware there is long standing concern in public health over the unequal distribution of health across social class groups, but our methods will pick up preferences over differences in achieved lifetime health and in periods of severe illness, both of which are expected to differ across social classes. Therefore, by implication, a premium for social class can be identified. Chapter 6 gives a simple illustration of how this can be done, where it is assumed that there is variation in life expectancy at birth across the social classes.

Study D explored the issue of concentration and dispersion. It has been pointed out that whilst people may prefer to spread out health benefits to a larger number of patients than to concentrate on a smaller number, if the benefit per person is below a certain threshold, then people may prefer to concentrate than to disperse. It therefore becomes important in devising priority setting questions that the differences in health will be large enough for people to recognise as something worth dispersing. Study D was a quasi-replication of a published study, using a set of new visual aids to take the respondents through the exercise. It identified 2.6 years as the point where the preference – over lifetime health and this is worth stressing – converts from concentration to dispersion. This finding was used to guide the design of trade-off exercises used in the remainder of the project so that the difference between two groups in any one scenario is at least 2.6 years.

CHAPTER 3: GENERAL DESIGN AND ANALYSIS ISSUES

Once we have determined the relevant attributes and the levels of those attributes, we need to consider how best to generate a set of equity weights for them. This chapter sets out the important design and analysis issues that need to be addressed if we are to elicit a robust set of equity weights from the preferences of the general public.

The first consideration concerns the way in which the trade-offs are to be specified. This is set out in 3.1, where we use a social welfare function (SWF) approach. The second consideration concerns what precisely the trade-offs are to be over (the ‘distribuendum’ as Dolan and Olsen, 2001, call it). This is discussed in 3.2, where we favour weightings over lifetime health. Section 3.3 then considers the kinds of questions needed for our purpose and sets out the main questions used in Phases 2 and 3. Section 3.4 presents the method of analysis used to parameterise the SWF.

3.1 THE TRADE-OFF (SOCIAL WELFARE) FUNCTION

In economics, the social welfare function is typically assumed to be a function of individual utilities, which are then weighted within the function to provide a trade-off in the utilities received by different beneficiaries (Atkinson, 1970; Little and Mirrlees, 1974; and Layard and Walters, 1994). Several studies have used a social welfare function in the area of health economics to model preferences and balance the competing demands of efficiency and equity (Dolan 1998, Dolan and Robinson, 2001, Abásolo and Tsuchiya, 2004). In health contexts, non-health outcomes for an individual are often disregarded and the focus is instead on health rather than utilities (Dolan, 1998). In one sense, this type of replacement is not unique to health economics, with replacements elsewhere of income for utility made where income inequalities are studied (e.g. Cowell et al, 1999). However, there is some ambiguity about whether health and income are taken to be proxies for utility or whether other arguments in the utility function are assumed to be constant.

Several functional forms have been suggested for the SWF, and these typically involve some form of concavity conferring value to a more even distribution of outcomes. Prominent amongst these has been the constant elasticity of substitution (CES; Dolan, 1988; Lindholm and Rosen, 1998) in which the health of two equal sized groups is assessed:

$$W = \left[\alpha v_1^{-r} + (1-\alpha) v_2^{-r}\right]^{-\frac{1}{r}}, \qquad \alpha \in [0,1],\ r \in [-1,\infty) \setminus 0$$

where: v_X is a judgement of the lifetime health of Group X,
α is the weight placed on the health of Group 1,
and r reflects the overall strength of inequality aversion.

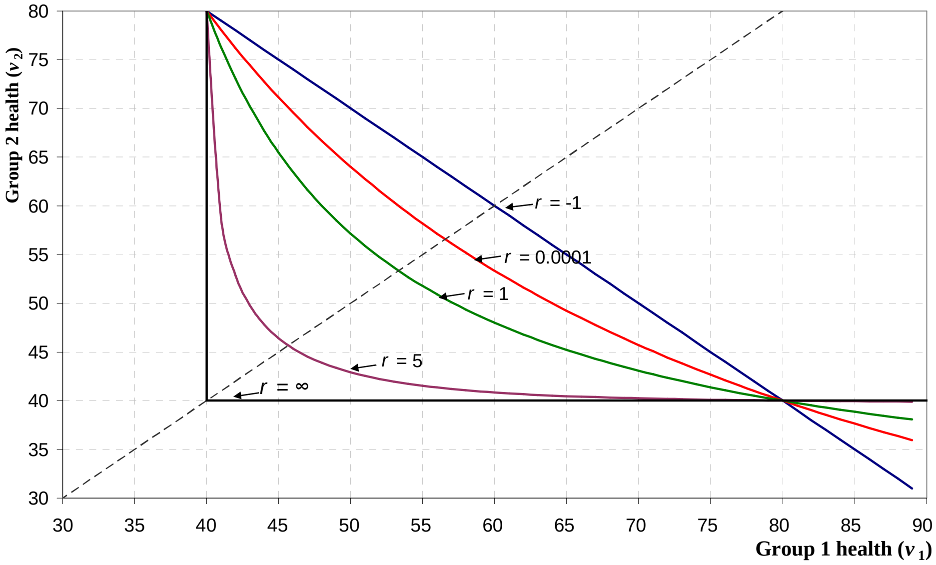
Other forms have been suggested, including the Cobb-Douglas form (Dolan, 1998), a special case of the CES, and the Atkinson form (Atkinson, 1970), used in Dolan and Robinson (2001). Furthermore, Abásolo and Tsuchiya (2004) have explored a SWF that can accommodate preferences that violate the monotonicity principle. The CES SWF is selected here because it is flexible enough to allow for a variety of different general types of preference.

The objective of the empirical study is to identify the inequality aversion parameter (r) and the relative weight (α), so that the marginal rate of substitution (MRS) for specific combinations of health can be calculated. The MRS in this context represents the relative social value of a marginal change in the social value of health to one group relative to the other, keeping the total level of social welfare constant. If the MRS is 1.5, that means that if the health of Group 1 deteriorates by 1 unit, then the health of Group 2 will need to improve by 1.5 units in order to maintain the current level of social welfare. This will then suggest that the marginal social value of the health of group 1 is 1.5 times that of group 2, indicating the relative values to be used in resource allocation decisions.

3.1.1 Inequality aversion: the effect of the *r* parameter

Through the r parameter, the CES SWF can encompass a variety of functional forms. One major restriction on preferences under the CES form is that they be *homothetic*. When homotheticity is satisfied, the trade-offs between different factors (the marginal rates of substitution) are unaffected by proportional increases in all variables. For identical individuals, we can set $\alpha = 0.5$. Suppose, for simplicity, that the QALY model describes the way that society values health, so that v is equal to the number of QALYs received. Here, Figure 3.1 is drawn with the health of Group 1 and 2 along the horizontal and vertical axes respectively. The graph shows combinations of health for the two Groups (v_1, v_2) that provide equal social welfare for various levels of inequality aversion. Starting from a point where Group 1 receives 80 QALYs and Group 2 receives 40 QALYs, the different iso-welfare curves all represent the same level of social welfare, albeit under different values of the r parameter for inequality aversion.

Figure 3.1: The effect of varying inequality aversion through a point.



For r = -1, and $\alpha = 0.5$, the function is a simple sum of the lifetime health of the two groups and no value is given to reducing inequality. Here, Groups 1 and 2 receiving 80 and 40 QALYs respectively is equivalent in social welfare terms to a case where both groups receive 60 QALYs, and:

$$W = 0.5v_1 + 0.5v_2 . \qquad (r = {}^{-}1)$$

As r rises, increasing value is given to equity. At $r = 1$, Groups 1 and 2 receiving 80 and 40 QALYs is equivalent in social welfare terms to a case where both groups receive 53.33 QALYs. Compared to the case where $r = -1$, this suggests a higher relative value to the health of the less well-off group.

At the extreme, only the group whose lifetime health is perceived to be worse is given any importance, so that a case where both groups receive 40 QALYs is equivalent to one in which Group 1 receives 80 QALYs (with Group 2 still receiving 40). This is the “maximin” SWF. Here:

$$\mathrm{W} = 0.5\min(v_1, v_2), \qquad (r \to \infty)$$

The value of inequality reduction is expected to differ according to the *relative* difference between v_1 and v_2. This means that if, other things equal, the size of the difference between v_1 and v_2 increases, we would expect that more value is placed on reducing this difference. If the size of v_1 and v_2 increases by the same amount, but the difference between them stays the same, then the relative size of the difference between the groups decreases and so we expect the value of inequality reduction to fall.

3.1.2 Non-health characteristics: the α parameter

The α parameter allows for the groups to be weighted differently even where there are no health differences between them e.g. because of different kinds or degrees of responsibility for health. The relative weight placed on a marginal improvement to the health of Group 1 is given by α. and the relative weight placed on the a marginal improvement to the health of Group 2 is $1 - \alpha$. The trade-off between the health of both groups (the marginal rate of substitution) is given by $\alpha/(1 - \alpha)$ along the 45^0 line.

Figure 3.2 is drawn with the health of Groups 1 and 2 along the horizontal and vertical axes, respectively. The graph shows how the iso-welfare curves for various values of α through $(v_1, v_2) = (80, 40)$ differ in the simplest case where there is no aversion to inequalities in health ($r = -1$). For $\alpha = 0.5$, the same value is placed on giving a marginal increase in health to either group (since $\alpha/(1 - \alpha) =0.5$). For $\alpha = 0.1$, nine times as much weight is placed on giving a marginal increase in health to those in Group 2 over those in Group 1 (since $\alpha/(1 - \alpha) =0.111$).

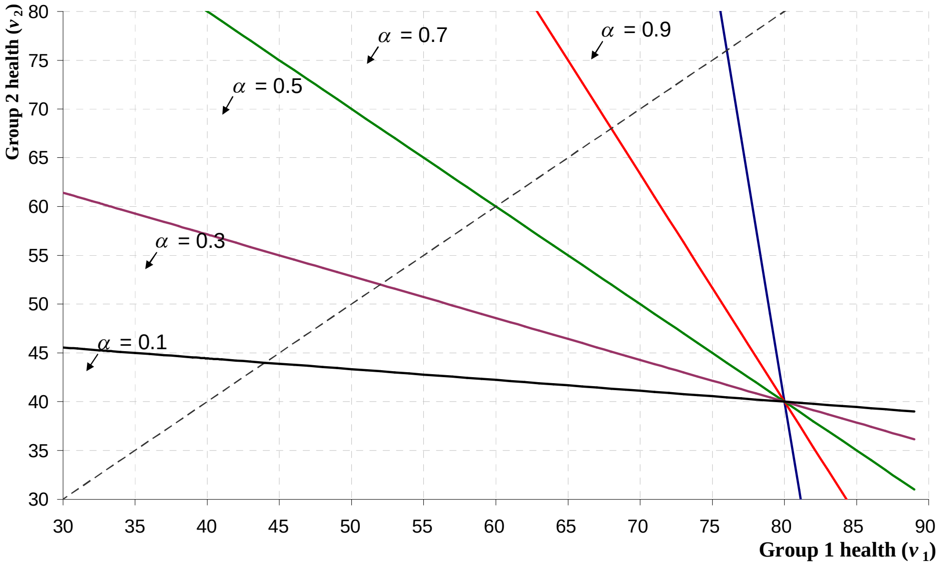
In general, the extra weight placed on a marginal increase in health for Group 1 is given by:

$$\frac{\mathrm{d}W}{\mathrm{d}v_1} = \alpha\, v_1^{-1-r}\left[\alpha v_1^{-r} + (1-\alpha)v_2^{-r}\right]^{-1-1/r}.$$

Whilst for Group 2 this figure equals:

$$\frac{\mathrm{d}W}{\mathrm{d}v_2} = (1-\alpha)v_2^{-1-r}\left[\alpha v_1^{-r} + (1-\alpha)v_2^{-r}\right]^{-1-1/r}.$$

Figure 3.2: The effect of varying α through a fixed initial point.



The trade-off for Group 2’s health in terms of Group 1’s health is equal to:

$$\frac{\mathrm{d}v_2}{\mathrm{d}v_1} = -\frac{\mathrm{d}W}{\mathrm{d}v_1} \bigg/ \frac{\mathrm{d}W}{\mathrm{d}v_2}$$

$$= -\frac{\alpha\, v_1^{-1-r}\left[\alpha v_1^{-r} + (1-\alpha) v_2^{-r}\right]^{-1-1/r}}{(1-\alpha) v_2^{-1-r}\left[\alpha v_1^{-r} + (1-\alpha) v_2^{-r}\right]^{-1-1/r}}$$

$$= -\frac{\alpha}{(1-\alpha)}\left(\frac{v_2}{v_1}\right)^{1+r}.$$

Take, for example, the case where $v_1 = 50$, $v_2 = 40$ and $r = 1$. Imagine that the trade-off between the healths of both groups is -1 (so that the same value is placed on a marginal increase in the health of both groups). Since the health of the two groups differ, inequality aversion (r) will give higher value to improving the health of group 2 over group 1. Therefore, in order for the health of the two groups to be valued equally, the relative weight (α) needs to cancel this out in favour of group 1, so that $\alpha > 1.0$. More formally:

$$-1 = -\frac{\alpha}{(1-\alpha)}\left(\frac{v_2}{v_1}\right)^{1+r} = -\frac{\alpha}{(1-\alpha)}\left(\frac{40}{50}\right)^{2} = -0.64\frac{\alpha}{(1-\alpha)}$$

$$1.56 = \frac{\alpha}{(1-\alpha)}.$$

This equation is solved where $\alpha = 0.61$. This suggests that in the absence of health differences, the marginal weight placed on the health of Group 2 is 1.56 times the marginal weight on the health of Group 1.

Given two (or more) states that are judged to be indifferent then for any set of SWF and lifetime health parameters we can estimate the social welfare attached to each state. The closer that these estimated values are within each pair (or set) of indifferent states, the more precise will be our parameter estimates.

In a case where there are no differences to base an unequal relative weight on, then $\alpha = 0.5$ and social welfare is a function of only one parameter (*r*) for inequality aversion. Let the suffix "FH" refer to full health, and "SH" refer to severe health (25% health), whilst "C" and "A" refer to child and adult timings. "Lifetime health judgements" (which represent the social value attached to profiles of health over a lifetime) can be written as:

$$v_i = y_{FHA} + FHC \cdot y_{FHC} + SHA \cdot y_{SHA} + SHC \cdot y_{SHC} \cdot,$$

where:

- *FHC* is the value of a year in full health whilst aged < 18,
- *SHC* is the value of a year in 25% health whilst aged < 18,
- *FHA* is the value of a year in full health whilst aged ≥ 18 (set to 1),
- *SHA* is the value of a year in 25% health whilst aged ≥ 18, and
- the *y* variables give the number of years spent in each health/time combination.

As we shall see more fully in Section 3.2.2 below, lifetime health judgements can be written as a function of two parameters, and so as few as three sets of pairs of indifference points are sufficient to find a social welfare function covering efficiency and inequality with an additional pair of indifference points necessary to find each weight for non-health characteristics. In practice, we can use more than this in order to have more confidence in the results. Of the three main attributes considered in the study, timing and severity of ill-health are interpreted as part of the definition of lifetime health judgements, and its effects are reflected in *r*. Condition cause/responsibility is considered a non-health characteristic and its effects are reflected in α.

3.1.3 Alternative formulations to the social welfare function

The framework of our analysis is motivated by something of a paradox: we cannot estimate the social value placed on reducing health inequalities without knowing how health is valued; and we cannot estimate the way health is valued (in a social setting, where more than one group is affected) without knowing how any differences between the groups affect these valuations. The major challenge of modelling societal preferences is in adopting a frame of analysis that is sufficiently flexible to allow both of these questions to be addressed within the same population. It is therefore preferable for both valuation and equity factors to be considered together. We believe our analysis is unique in estimating both the social valuation of lifetime health and the value given to more *equal* lifetime health together.

One elegant alternative to our selected analysis is provided by the rank-dependent model in Bleichrodt et al (2004) which considered the hypothetical lifetime QALYs of newborn. This model defined the QALY profile across society that assigns weight to a group according to their position in the health ranking across society (in QALY terms). This paper was reasonably conservative in its aims and considered health only between the ages of 10 and 40 years of age. It found an approximately linear function for the value of a QALY suggesting that each QALY in this range received the same weight and that those in worse health (according to ranking) received greater weight in the social value of a QALY. However, this analysis did not look at whether the

QALY model described preferences at a societal level nor could it consider the relative size of differences between groups.

Had our aim been to find an inequality measure rather than a SWF, our choice of tools may have been different. The CES function has been criticised as restrictive by those advocating inequality-based measures (Dutta, 2007), as has social welfare more generally (Bommier and Stecklov, 2002). An inequality measure does not consider a role for efficiency, and so is of limited usefulness in our current context, where the central aim of the project is to generate a set of weights that can be used in health care resource allocation decisions. Instead, an inequality measure attempts to identify choices that are “equally-unequal” rather than how good a particular outcome is when balancing efficiency and equity.

3.2 THE TRADE-OFF (OUTCOMES) SPACE

3.2.1 Gains or outcomes

Consistent with most of the work in this area, and in keeping with the design of our studies in Phase 1, ‘equity weights’ refer to any conscious departure from the assumption that all QALYs should be weighted equally. There is an issue, however, about whether we start with all QALY gains (i.e. the benefits from treatment) as being equally weighted or whether we consider final outcomes, which combine starting point and gains (Dolan and Olsen, 2001).

A gain framing takes the current position or status quo as the baseline and seeks to elicit trade-offs between maximisation gains and distributing those gains fairly. For example, suppose one individual could gain 5 QALYs whilst another individual could gain 15 QALYS, or else both could gain 10 QALYs. All else equal, an inequality averse respondent would prefer the second option. However, it is unlikely that all else would be equal. If the first person had already had 55 QALYs and the second person had already had 45 QALYs, then the first option will result in final outcomes of 60 QALYs each, and this may be judged to be the fairest option by an in inequality averse respondent.

It is possible to frame the questions in either way but, as the above example illustrates, the final outcomes space takes account of potentially relevant additional information e.g. in relation to overall health. It might be possible to break down the final outcomes into current position and gains but this would require yet more information for respondents to process. Since there is evidence that the greater the number of attributes presented simultaneously, the more likely individuals are to employ heuristics or shortcuts in making decisions (Payne et al, 1988; Lloyd 2003) rather than a substantive evaluation of the question.

When gains have arguably been the most salient consideration in the framing of the questions, there is some support for the notion that respondents are focussing on the final outcomes. If we consider the gains-space, then preferences should satisfy the Pareto Principle; that is, we should prefer to give benefits to one group if it does not imply a loss to the other group. However, we have found that up to 20% of respondents violate this basic principle (Dolan et al, 2002). This implies a backwards bending SWF (Abásolo and Tsuchiya, 2004) and it also provides evidence in support of a focus on final outcomes (which in these questions is more equal when the Pareto Principle is violated). Not violating the Principle may still be indicative of a focus on outcomes but without such a strong preference for reducing overall inequalities.

In any event, the use of gains is also problematic in that it requires the identification of a reference point. Where a reference point is identified there is evidence that individuals assess gains and losses differently, with losses given greater importance than equal sized gains (Kahneman and Tversky, 1979). That is, a loss of one QALY to one person must be compensated with more than a one QALY gain to another person. Even in cases where both options only involve gains from a status quo, there is evidence that one state will be taken as the reference point, and gains and losses calculated from this point (Dolan and Robinson, 2001). When considering whether a change is worthwhile, this asymmetric weighting of losses and gains makes it less likely that individuals will agree to any change. Moreover, the same consequences may be valued differently if what is regarded as the status quo changes. Finally, outcomes are the fundamental space within which the SWF is conceptualised. We therefore focus on the social value attached to profiles of health over a lifetime.

3.2.2 Societal valuations of lifetime health

The choice of outcome space as opposed to the gains space is also associated with the use of lifetime health experience as the relevant distribuendum. It is possible to discuss priority and equity in health care resource allocation based only on the current health of patients, with no reference to the lifetime health of these patients. For example, one may argue that if the cost-effectiveness of treatment is the same, patients who are currently suffering in severe health should be given higher priority over those patients who are only suffering a mild health problem. However, it may be the case that this severe suffering is only for a very brief duration, whereas the mild suffering is to last much longer; and if so, it is not obvious that the former patients should always get priority over the latter patients. This issue can be addressed by introducing an equal duration clause, so that the severe and mild suffering both last for the same duration. A more general approach to the issue is to take the lifetime perspective, which is what we do in this project.

In the simplest case, we could treat the social value of lifetime health as equal to the number of QALYs. The QALY gives no explicit weight to health at different life stages, as it is formed using only quality of life and duration. In contrast, societal weights may lead to different conclusions. We proceed by using the two attributes identified in Phase 1: childhood versus adult health and severe versus good health states.

The QALY takes the general form:

$$Q_i = \sum_{t=1}^{\infty} h_{it} \ ,$$

where h_{it} is the health-related quality of life of group i in time t. The lifetime health judgements based on the societal perspective may differ from how individuals value their own health in two respects: (i) the societal and individual perspectives may place a different value on ill-health; and (ii) the societal and individual perspectives may value ill-health at different ages in different ways.

Lifetime health judgments take the form:

$$v_i = \sum_{t=1}^{\infty} V(h_{it}, t),$$

where V is a weighting function based on health and timing (t) that increases in health ($\frac{dV}{dh_{it}} > 0$). Note that since our timing attribute takes two values, we can also write a formulation using the number of years spent in each health/timing combination. Here:

$$v_i = FHC \cdot y_{FHC} + SHC \cdot y_{SHC} + DC \cdot y_{DY}$$
$$+FHA \cdot y_{FHA} + SHA \cdot y_{SHA} + DA \cdot y_{DA},$$

where, in addition to the notations introduced earlier:

- DC is the value of death whilst aged < 18,
- DA is the value of death whilst dead aged ≥ 18,

We can scale these lifetime health judgements and do so by preserving the QALY values for death and full health for an adult (DA = 0, FHA = 1). Hence:

$$v_i = FHC \cdot y_{FHC} + SHC \cdot y_{SHC} + DC \cdot y_{DC} + y_{FHA} + SHA \cdot y_{SHA}. \qquad \text{(Eq.1)}$$

If we can also say that health and timing are multiplicatively separable:

$$v_i = \sum_{t=1}^{\infty} w(h_{it}) T(t) = t_0 \sum_{t=0}^{17} w(h_{it}) + \sum_{t=18}^{\infty} w(h_{it})$$

and $v_i = FHC \cdot y_{FHC} + SHC \cdot y_{SHC} + y_{FHA} + SHA \cdot y_{SHA}$,

$$v_i = y_{FHA} + FHC \cdot y_{FHC} + SHA \cdot y_{SHA} + (FHC \times SHA) \cdot y_{SHC}. \qquad \text{(Eq.2)}$$

Here, w is a lifetime health judgement, T is a weighting function on the timing of health, and h_{it} is the health of group i at time (or age) t.

As an example, consider the case where the following two states are regarded as equally good in terms of social welfare: In the first state, Groups 1 and 2 both experience 60 years of full health; in the second state, Group 1 experiences 65 years of full health, and Group 2 experiences 56 years of full health. Here, let us assume that $\alpha = 0.5$. Suppose, initially, that the value of health during childhood equals the value of health during adulthood (FHC=FHA=1), so that v_1=v_2=60 in the first state and $v_1 = 65$, $v_2 = 56$ in the second state. Here,

$$\left[0.5 \cdot (60)^{-r} + 0.5 \cdot (60)^{-r}\right]^{-\frac{1}{r}} = \left[0.5 \cdot (65)^{-r} + 0.5 \cdot (56)^{-r}\right]^{-\frac{1}{r}}.$$

Solving this numerically we find $r = 2.00$.

Suppose instead that the value of health during childhood is twice as high as the value of health during adulthood (FHC =2), so that now in the first state v_1=v_2=78 (since the first 18 years receive twice the weight of adult years) and $v_1 = 83$, $v_2 = 74$ in the second. Here,

$$\left[0.5 \cdot (78)^{-r} + 0.5 \cdot (78)^{-r}\right]^{-\frac{1}{r}} = \left[0.5 \cdot (83)^{-r} + 0.5 \cdot (74)^{-r}\right]^{-\frac{1}{r}}.$$

Solving this numerically we find $r = 2.89$. This illustrates how the parameters defining lifetime health judgements and inequality aversion are not independent from each other. Any observed level of aversion to in equality at the societal level is due to a combination of both the difference in lifetime health judgements and the significance of this difference to society.

Therefore, the inequality aversion parameter cannot be identified unless we know how big health differences are, their relative size, and the trade-offs society would make. Equally, as the preferences we observe are influenced by both inequality aversion and judgements about the value of health received, we cannot assess the size of the parameter defining FHC without knowing inequality aversion. We are, however, able to estimate these together.

In relation to the effect of non-health characteristics, consider the case where society is indifferent between the following states: In the first state, Groups 1 and 2 both experience 66 years of full health; in the second state, Group 1 experiences 60 years of full health, and Group 2 experiences 75 years of full health. If all years of full health receive the same weighting, then this suggests $r = 2.65$. If the health of Group 1 deteriorates by six years and the health of Group 2 improves by nine years, then it follows from that SWF that the level of social welfare will remain unchanged. Suppose that after personal responsibility characteristics are added into questions the second state needs to be changed to 62 years of full health to Group 1, 75 years of full health to Group 2 in order to achieve indifference between the two states. Society is now willing to give up only four years of Group 1’s health in order to get nine years of Group 2’s health. Group 1’s health is therefore given a greater value than before. In other words, a nine-year improvement to the health of Group 2 will now only make up for a four-year deterioration to the health of Group 1 as opposed to six, so the health of Group 1 is now given a greater value than before.

Here, adding personal responsibility characteristics have led to increased value on the health of Group 1; the best-fit α increases to 0.614 (versus 0.50 previously) since:

$$\left[0.614 \cdot (66)^{-2.65} + 0.386 \cdot (66)^{-2.65}\right]^{-\frac{1}{2.65}} = \left[0.614 \cdot (62)^{-2.65} + 0.386 \cdot (75)^{-2.65}\right]^{-\frac{1}{2.65}}$$

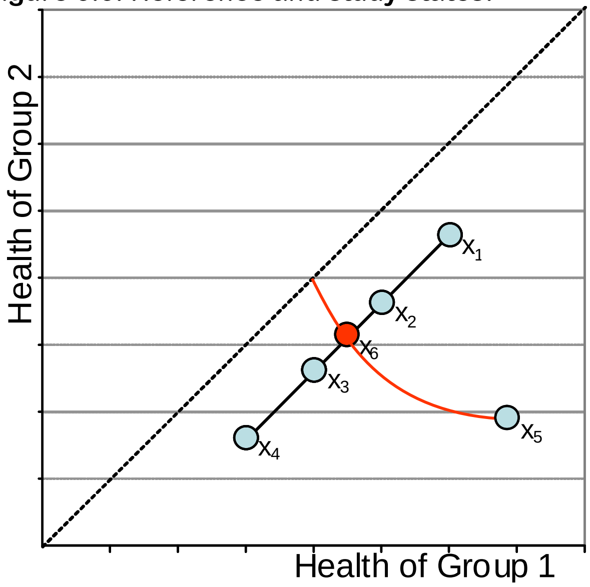
The inequality aversion parameter (r) does not influence preferences where the health of both groups is equal. The relative weight α is the marginal value of an improvement to the health of Group 1 relative to an improvement to the health of Group 2 where health is equal (0.614/0.386 = 1.59).

3.3 FINDING STATES WITH EQUAL SOCIAL WELFARE

Our general approach is motivated around identifying pairs of states with the same level of social welfare attached to them, where each state represents a different combination of lifetime health to two population groups. In order to find these states, our questions could use different methods. Our preferred method is a simple pairwise task that asks respondents which of a pair of states they prefer and may also allow for indifference.

Given the use of pairwise data, we need to be able to identify pairs of equally good states from a societal point of view. We do this by asking a series of questions that compare the same “study state” against four different “reference states”. For each study state we have a “Choice Set” of four independent choices. Within each Choice Set, we aggregate preferences to find a sixth point that has the same social welfare as the study state. Figure 3.3 shows the aim of this process, where reference states are labelled as x_1 to x_4, the study state is labelled as x_5 and the equivalent state is labelled as x_6.

Figure 3.3: Reference and study states.



Since health is higher for both groups in x_1 than in x_4, it is reasonable to assume that individuals will prefer x_1. In the same way, x_1 is preferred to x_2, x_2 is preferred to x_3, and x_3 is preferred to x_4. This is represented as: $x_1 \succ x_2 \succ x_3 \succ x_4$.

Suppose that (as in the example above) an individual prefers the study state (x_5) over the worst two reference states (x_3 and x_4) but prefers the best two reference states (x_1 and x_2) to the study state. For such an individual, $x_1 \succ x_2 \succ x_5 \succ x_3 \succ x_4$. This tells us where x_5 lies in the order of x_1 to x_4 for that individual.

In practice, individuals may make errors when making choices and might, for instance, indicate that they prefer x_5 to x_3 when the opposite is the case. This would mean that we might form the wrong preference ordering based on their individual data. In other cases, errors may prevent us from finding a preference ordering at all. For example, if an individual states that $x_5 \succ x_2$ and $x_5 \succ x_3$ then this suggests that $x_3 \succ x_2$, but we know that $x_3 \prec x_2$. In this case, we cannot form a coherent ordering between the individuals, and the assumption of *transitivity* is violated.

Whilst preferences at an individual level are “noisy”, we would expect these sorts of errors to balance out at an aggregate level. For this reason, we focus on preferences from all individuals together. That is, instead of finding where x_5 falls in a preference ordering for each individual, we instead consider where it falls in an *overall* ordering.

When we prefer one state to another, it is because the social welfare assessment we use (in that particular comparison) for the preferred state is higher than the social welfare assessment of the non-preferred state (in that particular comparison). If we had two identically-good states, then we would expect that half our sample would prefer one state and half would prefer the other state. $p(x_i, x_j)$ is defined as the proportion of the samply preferring x_i to x_j.

Thurstone (1927a, 1927b) outlined a simple method for scaling pairwise data that creates a cardinal scale, and we use the simplest (Case V) version here to scale social welfare judgements. This assumes that the social welfare assessments are independently and normally distributed with a common variance σ^2. The Thurstone score assumes that x_i is preferred to x_j when an individual perceives that x_i has the higher social welfare than x_j . If these assessments of social welfare are $W(x_i)$ and $W(x_j)$, then $p(x_i, x_j)$ is represents the proportion of our sample who perceive $W(x_i)$ as larger than $W(x_j)$. Under the assumptions above, this is a function of the underlying mean difference between the two states ($\overline{W_i} - \overline{W_j}$). The Thurstone score transforms each proportion into standard normal scores and finds the average score for all of the comparisons using a state. This average score is our (unscaled) estimate of $\overline{W_i}$, where we consider all those choices using x_i. We scale these scores so that $\overline{W_2}$ = 1 and $\overline{W_3}$ =0.

The Thurstone scores are based around a calculation of how often each state is preferred when it is compared to a state randomly chosen from all the possible states being compared (including itself). In this project, we consider a total of five states in each question. Here, there are 25 possible pairwise permutations, of which five see a state compared against itself (each state has a 50% chance of being preferred here). Of the remaining 20 permutations, we can infer the value of half of these from the remaining ten, since we know that one or the other must be preferred in each case. For the four reference states, the order of these states ($x_1 \succ x_2 \succ x_3 \succ x_4$) allows us to infer data since the monotonically superior state should be chosen in almost all cases. Six comparisons are provided in this way, leaving only four comparisons – x_5 versus x_1 to x_4 – to be identified in our survey work.

The score for the study state ($\overline{W_5}$) gives information about whether the study state is preferred in aggregate to each of the reference states. However, since these scaled scores are cardinally measurable social welfare values it also gives important information about how close the study state is (in social welfare terms) to these reference states.

Consider Figure 3.3. The lower is $\overline{W_5}$, the worse is the more unequal study state relative to the reference states, and so the less we have to move upwards from x_4 towards x_1 to find an equivalent point that has equal social welfare to the study state. The equivalent state (x_6) is defined as x_3 + $\overline{W_5}$ (x_2 - x_3). Where $\overline{W_5}$ =0, the equivalent state is x_3, whilst where $\overline{W_5}$ =1 the equivalent state is x_2. For $\overline{W_5}$ <0, the equivalent state is worse than x_3, and for $\overline{W_5}$ >0 it is better than x_2. It is possible that the equivalent state might be worse than the worst reference state or better than the best reference state in any comparison. The method will still allow the quantification of by how much better or worse the equivalent state (and thus the study state) is with respect to the reference states.

We can compute the scaled Thurstone scores using the aggregate preferences across all individuals, which then allow the construction of a set of equivalent states. These can then be used to estimate parameters in the SWF. Note, that whilst the values for the unscaled Thurstone scores are normally distributed, this is not the case for the scaled Thurstone scores, and hence also for the Thurstone-based equivalent state and the subsequent SWF parameters. As simple point estimates of uncertainty are of limited usefulness, we use bootstrapping to infer uncertainties. Bootstrapping assumes that the observed data are representative of the variation in the underlying population. A bootstrapping algorithm will select individuals at random and add them to a new (bootstrapped) dataset without removing them from the original dataset. This process continues until the bootstrapped dataset is of equal size to the original dataset. The analysis can be re-run on each new bootstrapped dataset and estimates obtained for equivalent states. Subsequent analyses can then also be re-run.

It may be possible to use other methods than the Thurstone score to infer equivalent states, including logistic and probit models. The four (direct) observations from each individual (x_5 versus x_1 to x_4) could be regressed on to dummies identifying the choice set. (Note that this method would not incorporate any information on the relationship between states x_1 to x_4, and may lose information contained in the study design.) The constant term in such a model could be used to identify the social welfare of the study state, with the coefficients on each dummy variable used to identity the social welfare of the other reference states. Overall, there is likely to be little difference to the predicted value of the study state if logistic or probit regressions were used in this way. (This is particularly the case for the probit model, since in both the probit and Thurstone analyses the main pieces of data are transformed using an inverse standard normal distribution.) Note, however, that the assessment of uncertainty may be different in each case, since the regressed model allows sampling from the distribution of coefficients estimated in the regression.

When re-sampling coefficients, there is a danger that the coefficients may be correlated. Bootstrapping, where each sample is constructed and analysed separately, seems less likely to suffer from this issue. If this issue is to be avoided, regression methods could be applied to bootstrapped samples. The main choice is between Thurstone scores that are almost instantaneous (using an Excel spreadsheet) and which use all the information contained in the study design, and regression methods that will typically take longer to compute and may be less efficient. Given this choice, we select the Thurstone method.

3.4 PARAMETERISING THE SOCIAL WELFARE FUNCTION

The Thurstone scores are used to infer an equivalent state (x_6) that is approximately indifferent to the study state (x_5). Given a set of *n* choice sets producing a pair of indifferent states (a study state plus equivalent state), define x_{i5} and x_{i6} as the *i*th pair of such states, where each is a vector of the time spent in each health/timing state (death whilst a child, severe ill health whilst a child, full health whilst a child, severe ill health as an adult, and full health as an adult). For simplicity, we re-define the function for lifetime health (consistent with the definition of Equation 1) as a function of the state considered and the parameters defining health for each group, *l*, that is:

$$v_l(x_{ij}; DC, SHC, FHC, SHA).$$

As our questions will always consider death after childhood, the value for *DC* will not be identified explicitly within the study, and so is excluded from the definition above.

Social welfare correspondingly becomes:

$$\begin{aligned}&\mathrm{W}(x_{ij}; r, \alpha, SHC, FHC, SHA)\\&= \left[\alpha\left(v_1(x_{ij}; SHC, FHC, SHA)\right)^{-r} + (1-\alpha)\left(v_2(x_{ij}; SHC, FHC, SHA)\right)^{-r}\right]^{-\frac{1}{r}}.\end{aligned}$$

In our estimated social welfare function, indifferent points should receive the same social welfare value (W), so that any difference between them can be interpreted as an error. The sum of squared differences errors across our data (*X*) is:

$$\begin{aligned}&\mathrm{E}(r, \alpha, SHC, FHC, SHA \mid X)\\&= \sum\nolimits_{i=1}^{n}\left(\mathrm{W}(x_{i6}; r, \alpha, SHC, FHC, SHA) - \mathrm{W}(x_{i5}; r, \alpha, SHC, FHC, SHA)\right)^2.\end{aligned}$$

We begin with the case where the groups are identical in all respects but health. Here, we expect $\alpha = 0.5$ to hold. From here, it would be preferable to simply minimise this error in order to find best fit estimates, whilst satisfying the multiplicative separability (so, also Equation 2). This would solve the problem:

$$\begin{aligned}&\text{minimise } \mathrm{E}(r, 0.5, SHC, FHC, SHA \mid X) \qquad \text{(Formulation 1)}\\&\quad\text{with respect to } r, SHC, FHC, SHA\\&\quad\text{where } SHC = FHC \times SHA.\end{aligned}$$

This problem must be solved numerically but will typically fail to converge to a solution. The issue here is that the size of the error in any choice is strongly related to the size of the parameters *r* and *FHC*. As above, as the value of *FHC* increases, so will the best-fit value for *r*. As this inequality aversion parameter increases, the difference in social welfare between any two states will tend to fall where the other parameters are the best-fit solutions for the level of *r*.

Consider the example given earlier with two equivalent states where in the first state Groups 1 and 2 both experience 60 years of full health; and in the second state Group 1 experiences 65 years of full health, and Group 2 experiences 56 years of full health. Suppose due to measurement error, the latter of these states was estimated to be 64.50 and 56 years in full health for Groups 1 and 2, respectively. Let us see what happens to the errors in social welfare as *r* increases.

Where the value of health during childhood equals the value of health during adulthood (*FHC*=*FHA*=1), v_1=v_2=60 in the first state and v_1 = 64.50, v_2 = 56 in the second state, with *r* = 2.00. The social welfare of the first state equals 60, and the social welfare of the second state equals 59.802. Where the value of health during childhood is valued twice as highly as adult health (*FHC* =2), the first state has v_1=v_2=78 and the second state has v_1 = 82.5, v_2 = 74, with *r* = 2.89 as before. Here, the social welfare of the first state equals 78.0000, and the social welfare of the second state equals 78.0238. The (absolute) error in the social welfare function is 0.1984 in the first case, and 0.1976. Whilst this difference is small, we have successfully reduced the error by increasing both *FHC* and *r*. This appears to hold generally, so that if we attempt to solve the problem as above *r* increases in every iteration, albeit at a very slow rate. The estimates for our parameters do not converge.

We therefore take a different approach where we do not solve for the best-fit value of *r* directly. Instead of trying to solve for *r* directly as above, we relax multiplicative separability and allow *SHC* to vary (so, use Equation 1). The value of the remaining parameters are found given a value for the inequality aversion, $\hat{r}$:

$$\text{minimise } \mathrm{E}(\hat{r}, SHC, FHC, SHA \mid X) \qquad \text{(Formulation 2)}$$
$$\text{with respect to } SHC, FHC, SHA.$$

We can therefore find parameters given a value of $\hat{r}$. Since the level of inequality aversion does not differ when solving this numerically, the convergence issues arising above are not relevant. We would still like to find a formulation where multiplicative separability (and hence Equation 2) holds, since there is a clear relationship between timing and health ($SHC = FHC \times SHA$) and childhood health states are anchored at death ($DC = FHC \times DA = 0$) and look for the value of $\hat{r}$ doing this.

Since we can identify values for all parameters given data and a value for $\hat{r}$, let us define the solutions to Formulation 2 above as $SHC(\hat{r} \mid X)$, $FHC(\hat{r} \mid X)$ and $SHA(\hat{r} \mid X)$. Now define the degree to which these best-fit estimates violate Equation 2, as:

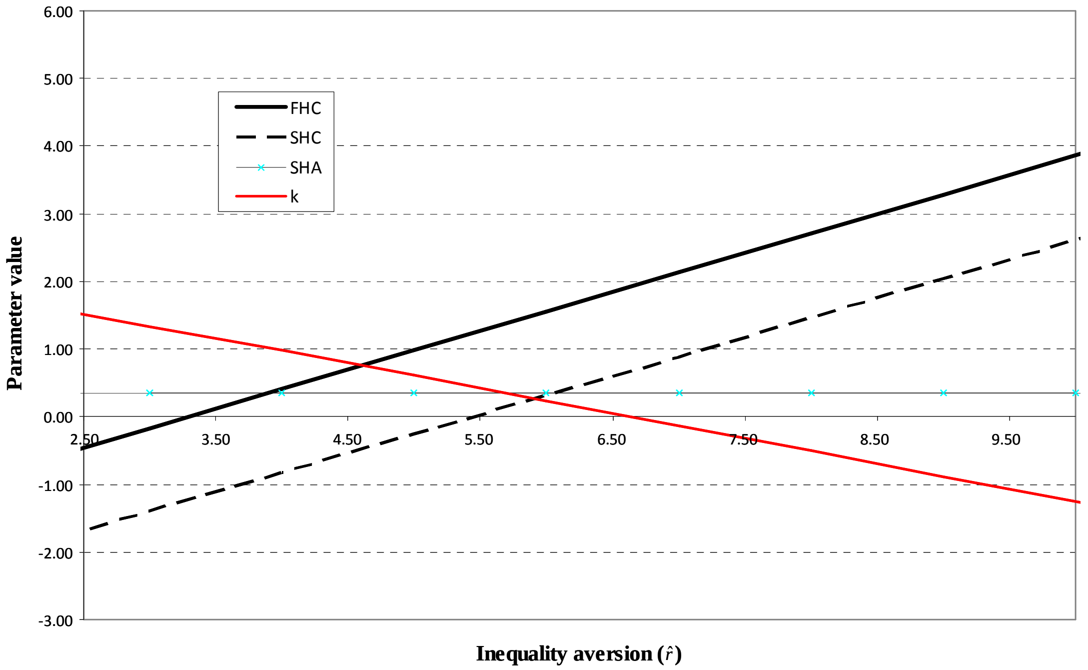
$$k(\hat{r} \mid X) = FHC(\hat{r} \mid X) \times SHA(\hat{r} \mid X) - SHC(\hat{r} \mid X).$$

Where $k(\hat{r} \mid X) = 0$, Equation 2 holds and we have a situation consistent with multiplicative separability between severity and timing in the societal valuation of health. Selecting this value of $\hat{r}$ and the parameter values (i.e. $SHC(\hat{r} \mid X)$, $FHC(\hat{r} \mid X)$, and $SHA(\hat{r} \mid X)$), we are able to parameterise the basic social welfare function where $\alpha = 0.5$.

This method does lead to convergent estimates for the social welfare function parameter *r*, and the parameters defining health judgments the (v_i). Figure 3.4 gives a example of how these parameter values differ with $\hat{r}$ for sample questionnaire data. All curves here are a typical representation of how the parameter values differ as r is varied (i.e. for moves along the horizontal axis). (As above, we selecting $\hat{r}$ where k=0, and hence where the *k* curve it crosses the horizontal axis.) Note that the value of severe health as an adult is nearly constant with respect to inequality aversion, whilst the values for childhood health are approximately linear. This property also aids convergence.

Within the general form of the SWF, the parameter α is used to capture the degree to which groups are treated differently for non-health reasons. The general method in finding values for this is to compare cases where questions include a particular characteristic – for example, a difference in the cause of a condition – with others whether the question do not.

Figure 3.4: Lifetime health parameters by inequality aversion



Bootstrapping is used (as above) to provide more sets of indifferent states on which these analyses can be run. By re-running the analyses on these new sets of states, we estimate the uncertainty in the parameters we find. We are not aware of any other way of estimating these uncertainties as the parameters are found using numerical methods that will not necessarily follow a specific distribution.

3.5 CONCLUSION

This chapter has addressed general issues around the design of the survey and the analyses of the results. This project employs the SWF approach, using the CES specification, based on the lifetime health experiences of different groups of people. The CES SWF involves two key parameters: one is the inequality aversion parameter (r), which represents the degree of trade-off between overall level of health and a more equal distribution of this overall health across different population groups when they differ only in terms of health; the other is the relative weight (α) for non-health characteristics, such as responsibility for ill health. In terms of the distribuendum, the CES SWF used in this project is defined in terms of lifetime health of people in different population subgroups, in the health outcome space.

The empirical survey presents respondents with a series of pairwise choices. Each choice is between two states involving two population groups, and is described in terms of the lifetime health and the characteristics of the people in each group. The objective of the survey is to elicit preferences from the members of the public so that the main parameters of the SWF can be identified. Once these are estimated, then the MRS for different states can be calculated.

CHAPTER 4: GENERATING STABLE PREFERENCES

If preferences are to be used to help guide policy, it is important that these preferences are stable. Phase 2 of the project sought to compare the effects of deliberation on preferences. The two designs used in Phase 2 primarily differ in the use of group discussions on related questions and an opportunity for personal reflection prior to a face to face interview in which the main preference elicitation task takes place.

4.1 FORMS OF DELIBERATION

4.1.1 Attitudinal questions

It has been suggested that researchers should encourage deliberation of pertinent issues before preferences are elicited in an attempt to facilitate more stable responses (Clarke, 1999). Research also suggests that more informed decisions are made when individuals systematically evaluate all the available information in accordance with their existing beliefs (Bekker and Walker 2001). Informed, stable preferences of this type would arguably provide valuable information for policy makers.

One way of facilitating deliberation may be to encourage respondents to examine their own attitudes using a series of questions (details below) prior to receiving the main preference elicitation task. These involve a range of questions about health and political beliefs, including the degree to which individuals feel priority should be given to named groups. By considering alternative “deserving” groups, it is implicit that some choice must be made between groups. The process of considering alternative groups is intended to help clarify the respondents’ thoughts on the topic, and so lead to more stable preferences at the later stage. General attitudinal questions have been found to reduce variability in a previous elicitation studies in Ratcliffe et al (2007).

The “beliefs questionnaire” elicits the respondents’ health and political beliefs and their willingness to prioritise NHS resources towards specific groups. The first 20 health and political belief questions cover attitudinal questions on general social issues and social attitude questions on health. In both cases, respondents were asked to state how much they agree or disagree with the statement listed. These statements are taken from the British Household Panel Survey (Wave 14), and attitudinal questions used in Ratcliffe et al (2007). The topics in the following 10 questions on willingness to prioritise NHS resources were based on the Phase 1 studies A and B. As the majority of the questions used here were directly taken or slightly modified from existing questions, it was not felt that the wording of these questions needed to be explicitly piloted prior to Phase 2.

4.1.2 Discussion groups

Economists generally assume that experience provides a feedback mechanism that allows individuals to decide upon the value of particular goods and services. In more abstract settings, preferences cannot be based on direct experience since no practical forum exists to gain it. In this case, whilst it is not possible for experience to provide feedback on values, it is possible that some form of discussion and deliberation may be able to. If such discussion has the effect of improving the information on which a decision is made, then it arguably provides better preferences. Our earlier work (e.g. Dolan et al, 1999) and our proposal was generally in favour of discussion.

However, we became more reticent after further discussion with the psychologists in our team. The phenomenon of group polarisation is well-established within psychology (Moscovici and Zavalloni, 1969), whereby group interaction tends to exaggerate attitudes within a group. For example, “risky shifts” occur where a group’s decisions are riskier than the average of individual’s own decisions in isolation (Stoner, 1961). Similar types of changes see mock juries made up of authoritarians imposing more severe penalties than they would have recommended when deciding as individuals, just as less authoritarian mock juries give more lenient penalties (Bray and Noble, 1978). Where such mechanisms occur, the normative place of discussion and subsequent deliberation is more questionable.

In general terms, two important classes of theory are important for us to consider here: social comparison theories and persuasive argumentation theories. Isenberg (1986) defines social comparison theories (SCT) as those assuming that individuals “are constantly motivated both to perceive and to present themselves in a socially desirable light. In order to do this, an individual must be continually processing information about how other people present themselves, and adjusting his or her own self-presentation accordingly.” Under SCT individuals “preferences” represent not only the underlying values of an individual but also an element of gaming –the intentional misrepresentation of her preferences to improve their social acceptability.

Whether group polarisation provides better choices under SCT may depend on the details of the situation. An individual may have an incentive to present themselves as generally typical and differing by a little to signal the direction in which an individual would like society to move. If individuals typically misjudge the average group value then prior to group communication they would self-present some distance from society’s true value. As individuals gain more experience and “deliberate” (or even if they are merely exposed to other’s values) they update their expectations and move closer to true societal values. In such a case, deliberation may be positive in the sense that it provides better information about society’s values. However, suppose instead that individuals define themselves against their expectation of society’s values. If a group of like-minded individuals react with each other they will tend to become increasingly extreme in their positions, since each defines their position against the average of those they observe, which itself is becoming more extreme.

Persuasive argumentation theory (PAT) “holds that an individual’s choice or position on an issue is a function of the number and persuasiveness of pro and con arguments that that person recalls from memory when formulating his or her own position” (Isenberg, 1986). It is normally suggested that persuasiveness is based on the novelty and validity of an argument, with novel and valid arguments trivially preferred to established or invalid ones. Sub-optimal choices occur where individuals do not have access to all arguments and also where some arguments are novel, since this will tend to bias conclusions towards newly-discovered arguments. Group discussion in PAT could either improve or worsen the quality of deliberations, depending on whether the bias from ignored arguments or the bias from novel arguments dominates. As the time for deliberation increases, the novelty of new arguments will diminish, and PAT will view a discussion and (sufficient) deliberation process as beneficial.

If the deliberative effects found in the experiment were due to improvements in knowledge about the values others hold then it is arguable that preferences after discussion are superior. If instead changes were due to bandwagon effects, then "colder" preferences are probably preferable. Whilst economics assumes that we reflect and learn from our own choices to produce more rational conclusions, this assumption is not necessarily true. As a phenomenon, group polarisation provides an interesting puzzle, and the differing models of human choice raise questions as to whether deliberative preferences (whether SCT or PAT based) are automatically preferable to "cold" preferences.

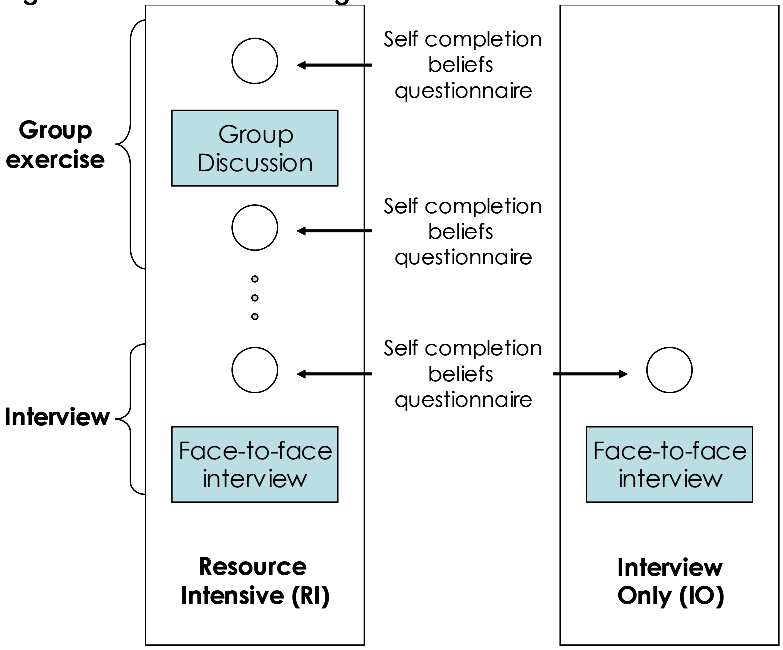
4.2 TESTING FOR THE EFFECTS OF DISCUSSION

4.2.1 Methods

4.2.1.1 Overall design

Against this background, and with resource considerations firmly in mind, we decided to use attitude and beliefs questions in a self-completion questionnaire as a pre-cursor to all the preference elicitation tasks but to test the feasibility and effects of group discussions. Phase 2 of the project explores the differences between a more intensive, more costly design, involving a group discussion and a face-to-face preference elicitation interview (referred to as Resource Intensive, or RI), and a less intensive and less costly design, involving attitudinal questions and a face-to-face preference elicitation interview (Interview Only, or IO). Details of these designs are given in Figure 4.1.

Figure 4.1: Stages in the RI and IO designs.



The RI design begins with a group exercise where respondents are asked to complete a simple self-completion questionnaire individually, which will ask them a set of general attitudinal questions relating to health and political beliefs. This is followed by a discussion session covering the relevant issues and arguments. The discussions will focus on the reasons for particular preferences, and whether or not respondents are happy with the policy implications of their preferences. This discussion session was of a similar form to those used in the Phase 1 studies A and B. The self-completion questionnaire is administered again at the end of the session and a time arranged for an interview one or two weeks later at the respondent's home. This interview begins with a third administration of the self-completion questionnaire. The main preference elicitation task then follows.

The IO design excludes the pre-interview portion of the design, and comprises of the self-completion questionnaire and the main preference elicitation interview. These were similar to those given to the RI group, with the only difference being a slightly expanded introduction to the interview. The self-completion beliefs questionnaire also encourages a baseline level of deliberation in the IO design, since we wish all our respondents to answer the preference elicitation questions with at least some familiarity with the issues to be raised. For the RI group, this questionnaire also provides an indication of the effect of discussion and deliberation as each respondent completes it three times. A copy of the prompts and interview used in the RI variants are given in Appendix 4.

The discussion points use the same questions as used in the Phase 1 studies A and B and consider the explicit gains to particular groups differentiated by a variety of factors. These factors include age, the cause, and the severity of their conditions. As in the Phase 1 studies, the respondents were asked to discuss whether they would be willing to accept a lower QALY gain overall in order to prioritise one group over the other. As with the self-completion beliefs questionnaire, it was not felt necessary to pilot the wording of the questionnaire as it follows very closely to that used successfully in the Phase 1 studies.

The preference elicitation task uses a series of choices that concern the health of two equal-sized groups. In essence, "a study state" is compared to four "reference states" and the relative values of these five states are identified through the implied ranking of these states relative to each other (see Section 3).

4.2.1.2 The questions

The preference elicitation task used in both Phase 2 and Phase 3 is comprised of 16 choice sets in six questions, each composed of four choices between two states at a time. Table 4.1 summarises these choice sets and their purposes. Choice Sets 1.1-1.4 involve choices where all life years are lived in 100% health, after which all those in the groups die. There are no differences between the groups in the type of condition experienced, the rarity of the condition, or the numbers in each group. If we also assume that the health of adults and children are valued equally, then we expect a SWF with a CES functional form where lifetime health is measured in QALYs. As such, we do not need to worry about the relationship between lifetime health and inequality aversion here. In this case, violations of the CES functional form would suggest either that the CES form is inappropriate or that the QALY does not describe lifetime health. Responses to this question can also be used to find a value for the inequality aversion (r) parameter within the SWF, and provide an opportunity to gain familiarity with the question format.

Table 4.1: The structure of the preference elicitation task

Question	Choice Sets	Illness	Data used to derive:
1	1.1-1.4	No	Inequality aversion (standard QALY model assumed to hold)
2	2.1-2.4	Yes	Lifetime health judgements (social value of timing of ill health) inequality aversion (where QALY model does not hold).
3 & 4	3.1-3.2, 4.1-4.2	Yes	Effect of condition cause/responsibility (abstract descriptions) (with 2.1-2.4)
5	5.1-5.2	Yes	Effect of condition cause/responsibility (labels) (with 2.1-2.4, 3.1, 4.1)
6	6.1-6.2	Yes	Effect of condition rarity (with 1.2)

Choice Sets	Description of Group 1	Description of Group 2
1.1-2.4	Groups 1 and 2 are identical except in the health that they experience	
3.1-3.2	Those in Group 1 experience an illness that is not the result of their lifestyles but is instead caused by errors within the NHS.	Those in Group 2 experience an illness that is caused by a combination of factors including poverty, genetics, pollution, and the patients' lifestyles.
5.1	Those in Group 1 experience an illness that is due to MRSA ("superbug") infections picked up after NHS operations.	Those in Group 2 experience an illness that is caused by obesity.
4.1-4.2	Those in Group 1 experience an illness that is not the result of their lifestyles but is instead caused by errors within the NHS.	Those in Group 2 experience an illness that is caused by a combination of factors including poverty, genetics and pollution, but is not caused by patients' lifestyles or by NHS error.
5.2	Those in Group 1 experience an illness that is due to MRSA ("superbug") infections picked up after NHS operations.**[Both Phases]**	Those in Group 3 experience an illness that is due to workplace exposure to hazardous substances (e.g. asbestosis). **[Phase 2]** Those in Group 3 experience an illness that is due to a genetic condition that affects the health of people in middle-age. **[Phase 3]**
6.1	Those in Group 1 and Group 2 both experience illnesses that are caused by a combination of factors including poverty, genetics and pollution, but are not caused by patients' lifestyles or by NHS error. There are equal numbers of patients in both groups.	
6.2	Those in Group 3 and Group 4 both experience illnesses that are caused by a combination of factors including poverty, genetics and pollution, but are not caused by patients' lifestyles or by NHS error. The illness affecting those in Group 3 is extremely rare, and the illness affecting those in Group 4 is slightly more common. Your choice will affect an equal number of patients in Groups 3 and 4.	

Choice Sets 2.1-2.4 again involve choices in which there are no differences between groups on rarity and condition cause/responsibility grounds. In these choices, individuals may experience a period of ill health at 25% quality of life. These questions are used to find weights given for childhood versus adult health, 25% health versus 100% health, and a value of r. The next choice sets introduce a condition cause/responsibility dimension into choices that are otherwise identical to those in Choice Sets 2.1 and 2.3. Choice Sets 3.1 and 3.2 compare a Group 1 with NHS caused illnesses with a Group 2 that has illnesses partly caused by the patient's lifestyle. Choice Sets 4.1 and 4.2 compares a Group 1 with NHS caused illnesses with a Group 2 that has illnesses not caused by the patient's lifestyle. These questions are used to find global (α) weights representing the effect of condition cause.

Choice Sets 5.1 and 5.2 repeat Sets 3.1 and 4.1 but name a more specific cause of illness rather than a general description of the type of illness. In place of NHS caused illnesses we have MRSA (Methicillin-Resistant Streptococcus Aureus) infections; non-NHS, partial patient condition causes are instead labelled "obesity"; and non-patient, non-NHS illnesses are labelled as "workplace exposure to hazards" in Phase 2. (In Phase 3 "genetic disorder" is used). These questions are used to find the effect of providing further information about illnesses. As the description of the causes are the only differences between Choice Sets 3.1 and 5.1 (and 4.1 and 5.2), we can tell whether preferences differ in the more-abstract categorical (Choice Set 3.1/4.1) and the less-abstract labelled case (Choice Set 5.1/5.2).

Choice Sets 6.1 and 6.2 are largely stand-alone and assess the effect of rarity of diseases on choices. In Choice Set 6.1, both groups suffer from an equally common condition and in Choice Set 6.2 one group suffers from an extremely rare condition. These labels are used to consider the case of ultra-orphan conditions. No additional information was given to respondents about the policy context since extreme rarity is only one characteristic of such conditions. We were also concerned that adding a lengthy exposition of the policy question in this one case would lead respondents to infer particular significance to rarity. Note that respondents were instructed that the numbers affected in each group remain the same in all four cases.

4.2.1.3 Piloting

4.2.1.3.1 Resource intensive design

The aim here was to conduct one discussion group with about six respondents. Potential respondents were approached in Sheffield city centre. If they agreed to take part, they were given the respondent information sheet, consent form and university map detailing the location for the group exercise. Participants were also mailed a letter of confirmation. Eleven individuals agreed to participate, of which five were absent and one was unable to attend due to illness. Follow up face-to-face interviews were obtained with all five respondents, each receiving £15 for their participation. Participant and interviewer feedback was recorded following the interview.

In the version of the preference elicitation task used in this piloting, the four reference states were outlined prior to the study state with the best of the states at the top, and the worst of the states at the bottom. The respondent was asked whether they agreed with this ordering, so that x_1 was compared to x_2, x_2 was compared to x_3, and so on.

Next, the study state (x_5) was compared to each of the reference states (x_1 to x_4) in order. A preference order between the study and reference states can be found as soon as an individual prefers the study state in a choice (or was indifferent between the study and reference state).

Preferences from all five respondents were coherent in the sense that a transitive ordering exists, but this is unsurprising in light of the method used. All interviews lasted between 30 and 80 minutes. Our main concern was that the format of the elicitation task may be too complex for the interviewer to administer. It was also felt that the question format was too restrictive in encouraging transitive behaviour, and that this may remove some of the natural variability that should be expressed in the Thurstone index. The pairwise elicitation task was simplified in light of these concerns. Following this first piloting stage, each choice set is now comprised of four stand-alone questions comparing a single study state against four reference states. Whilst the reference states continue to have a monotonic ordering, the preferences between them are no longer tested directly. This also allowed us to change the order in which the reference states were presented, although at this stage we retained the best-to-worst order for the reference states in each choice set.

4.2.1.3.2 Interview only design

Five respondents were interviewed by an interviewer 'cold calling' at people's homes in Sheffield. All respondents were given £5 for participating in the research. As in the RI design, the IO design begins with a self-completion beliefs questionnaire eliciting both health and political beliefs and the respondent's willingness to prioritise NHS resources towards specific groups. Following these self-completion beliefs questions, respondents answered the preference elicitation task. At the end of the interview, background characteristics, respondent feedback and interviewer perceptions were recorded. As expected, the responses from these individuals typically included a higher level of intransitive choice. Whilst the choice task is complex, we believe that this format could not be easily simplified and that the majority of respondents appeared to be capable of answering the questions posed, with our interviewers able to identify questionable data prior to analysis.

4.2.1.4 Format of questions

As both the RI and IO versions performed acceptably in piloting, few changes were made subsequent to the piloting process. In each choice, respondents were asked which of two states they preferred and indifference between the states was accepted. Each choice is prefaced by a text box giving information about the choice context followed by the choice itself. As an illustration, Figure 4.2 displays one of the choices used in Choice Set 5.1. Quality of life is explained to respondents by stating that, for instance "A year in full health is four times as valuable as a year in 25% health to the person living it".

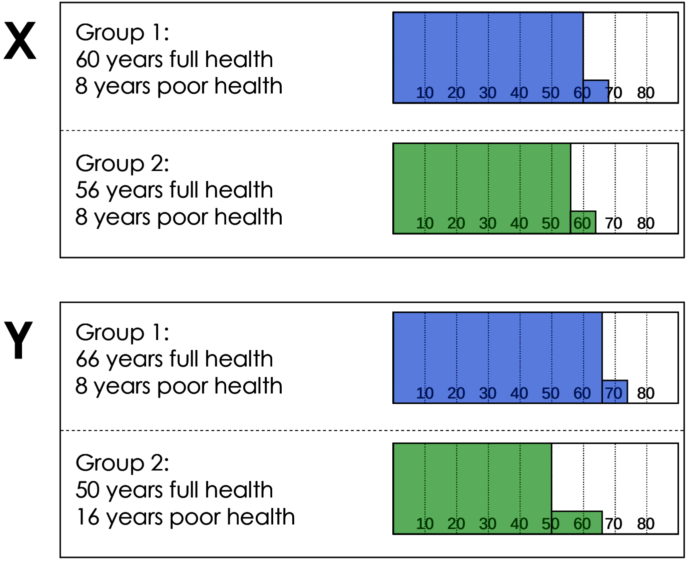
Figure 4.2: Example of choice context box and choice diagrams.

About the groups:

Those in Group 1 experience an illness that is due to MRSA (“superbug”) infections picked up after NHS operations.

Those in Group 2 experience an illness that is caused by obesity.

The groups are similar in all other ways.



4.2.1.5 Recruitment

It is difficult to calculate the necessary sample size for comparing the designs since Thurstone calculations used to analyse the preference elicitation tasks are based on bootstrapping (where no calculations are possible) whilst the factor analysis used to analyse the self-completion beliefs questionnaire produces a composite variable in a form that is not observable until the analysis is underway. We initially attempted to recruit 144 individuals for both the IO and RI arms of the experiment as this was our intention in the proposal. Six areas were sampled (Barnsley, Cumbria, Dronfield, Huddersfield, Rotherham and Sheffield) at Phase 2. A total of 3342 households received information sheets relating to the project, but not all of these households were contacted by interviewers when seeking recruits to the project. Representativeness is not essential here, although it is important that in Phase 3 our final sample reflects a range of age, sex, and educational characteristics. Eight interviewers were briefed to undertake about 18 interviews of each variant.

All households were sent an ‘interview only’ respondent information sheet with a covering letter detailing the study and explaining the possibility of a RI option. Respondents would be paid £5 for attending the IO and £15 for the RI variant. When meeting the potential interviewees, interviewers would offer the choice of participation in the RI variant. All interviewers aimed to recruit for both variants on the doorstep. Interviews for the IO variant were conducted straight away or at a mutually agreed time. Those who agreed in principle to participate in the RI variant were then telephoned and offered days/times to participate and then sent a letter confirming the time, date and venue. In practice, this resulted in some individuals being contacted up to six times before confirming attendance, and even here some still dropped out.

We initially tried to accommodate people from their availability and arrange groups around them. When this proved to be difficult, we then told people when the groups were and asked them to attend. There seemed to be little difference in level of response from the two options. Groups were offered with over two weeks' notice and also shorter notice periods, but again there seemed to be little difference between the two options: those with a longer lead time seemed to have forgotten their agreement to participate and those with a shorter lead time were not able to attend at short notice. Recruitment to the RI design proved to be logistically difficult, and of the target of 144 respondents only 56 were recruited in the allotted time. In contrast, we over-sampled within the IO group, with 232 respondents interviewed.

4.2.1.6 Analysis

Thurstone scores were calculated for the sixteen questions in the preference elicitation task, and an estimate of the variability in each was computed using bootstrapping. As the preference elicitation task was run once in each design, this provides results that can be compared using a two-sample z-test with n_1=n_2=1. (A significance level is set at 5% for all significance tests. Whilst we conduct multiple tests in many of these stages, the use of the Bonferroni or similar methods of adjustment are not warranted.) Any differences between the RI and IO designs could suggest that discussion and deliberation affects preferences but this would provide weak evidence only, as other factors may be responsible. In particular, given that randomisation was not feasible, selection effects or random chance could be responsible, the latter of which is accounted for in the test of Thurstone scores.

The tests on the (unscaled) Thurstone scores are also unable to consider the separate impacts of group discussion and personal deliberation because the preference elicitation task was used only once for the RI group after both discussion and deliberation had occurred. In contrast, data from the self-completion questionnaire was found at multiple timings, and includes prioritisation questions that will allow the estimation of separate effects. We therefore analyse a combined data set including each use of the self-completion beliefs questionnaire – one time in IO and three times in RI.

Therefore, in addition to testing for differences from the preference elicitation task, we also consider the following questions:

1) Are background characteristics the same across the designs?
2) Are there differences in the self-completion health and political belief questions (over and above the effect of any differences in background characteristics)?
3) Does design affect the willingness to prioritise some groups (over and above the effects of background health and political beliefs)?

We collected several types of background characteristics including, age, gender, ethnic identity, marital status, housing status, employment status, education status, life satisfaction ratings, health ratings (excellent, good, fair, poor, very poor), and whether the respondent has private health insurance. We test the first question by analysing whether differences exist between the two samples.

The 20 attitudinal questions regarding health and political belief questions were first analysed using factor analysis techniques in SPSS. The extraction technique was principal component analysis and Varimax rotation was used to construct the component matrix. The number of variables was defined using a Scree plot, with cut-off at first flattening of the curve. Kaiser-Meyer-Olkin measure of sampling adequacy and Bartlett’s test of sphericity were both run to test for the appropriateness of factor analysis.

The resultant factors were computed for each individual and regressed against background variables and dummy variables indicating the design/timing of the questionnaire (RI pre-discussion, RI post-discussion, RI pre-interview, and IO pre-interview). The baseline comparison at first contact is between RI pre-discussion and IO pre-interview, whilst the relevant comparison for equality of data in the main elicitation task is RI pre-interview and IO pre-interview. Comparison of the three RI timings provides an indication of whether, and how, deliberation affects health care beliefs over and above the effects of background characteristics, and hence addresses the second issue raised above.

Factor analysis was also applied to the 10 prioritisation questions of the self-completion questionnaire using the same methods as used for the first 20 belief questions. These questions were also analysed using a regression of each prioritisation factor against the factors found in the analysis of the earlier (belief) questions, design/timing dummies, and background variables.

4.2.2 Results

4.2.2.1 Differences in respondent characteristics

Table 4.2 gives details on the background of each sample. The IO design appears to be more representative against Wave 14 of the British Household Panel Survey (BHPS, 2006). Both groups over-sample females, the retired, students and those who have never been married. Those in the RI design tend to be slightly healthier than the BHPS sample; those in the IO design marginally more ill. The IO group appears to be closer in the housing, health insurance and the number of disabled included. Note that our background variables consider age bands, whilst the BHPS uses continuous ages and this makes direct comparison difficult. (Like the BHPS, we undersample ethnic minorities.)

There are several significant differences between the RI and IO samples, which typically reflect socioeconomic differences. In particular, whilst 91% of the RI sample owned their own house (outright or with a mortgage), only 75% of the IO sample did (p=0.007, Fisher exact test). University education was more common within the RI respondents (35% vs. 19%), as were the proportions of self-employed (13% vs. 6%) and retired (45% vs. 36%) people. No significant differences appear on general life satisfaction, gender, age, ethnicity, or marital status.

Whilst not significant, the RI group had a third as many people self-declaring disabled as the IO group (4% vs. 12%). The IO group had higher use of GP services in the last three months (47% no visits, 49% 1 or 2 visits, 5% 3 or more visits) versus the RI group (41%, 38%, 21%), and these differences were significant (p = 0.011). However, there were no significant differences in either the distributions of self-declared health status (excellent, good, fair, poor, very poor), or in the rates of health insurance (22% vs. 15%). As such, whilst there may be some underlying health differences between the samples, the majority of the differences found here were socioeconomic in nature.

Table 4.2 Background of the sample

Sample size		RI design % (n=56)	IO design % (n=232)	Significant difference
Gender:	Female	55	63	0.361
Age:	18-29	3	12	0.302
	30-39	18	16	
	40-49	16	18	
	50-59	14	18	
	60-69	29	18	
	70+	20	19	
Ethnicity	White	98	99	0.225
Marital status:	Married/cohabiting	77	63	0.170
	Separated or divorced	5	14	
	Widowed	5	10	
	Never married	13	13	
Employment status:	Self-employed	13	6	**0.033**
	Employed	36	39	
	Retired	45	36	
Education:	Comprehensive	20	25	**0.009**
	Grammar (all)	15	7	
	University	35	19	
House ownership:	Owned/mortgage	91	75	**0.007**
Disabled?	Yes	4	12	0.086
Health status:	Excellent	36	25	0.057
	Good	48	38	
	Fair	11	23	
	Poor	4	10	
	Very Poor	2	5	
Medical insurance:	Yes	22	15	0.697

4.2.2.2 Differences in attitudes and beliefs

Unlike background differences, the health and political beliefs of each individual may change *within* the RI design. The median responses for set of questionnaire observations are given in Table 4.3. In 14 of the 20 questions the median response remains the same across all three RI timings (pre-discussion, post- discussion, and pre-interview). In two of the remaining six questions, the median response changed during the discussion period. In the last four, this change occurred within the deliberation period post-discussion and pre-interview. None of the changes between timings are significant on any one variable (using mean responses). Given the small sample size this is possibly inevitable.

Table 4.3: Health and political belief questions by observation set

	Median			
	RI pre-discussion	RI post-discussion	RI pre-interview	IO pre-interview
"Government should redistribute income from the better-off to those who are less well-off."	3	2	2	3
"Income tax should be increased to improve the running of schools and hospitals."	3	3	3	3
"Private enterprise is the best way to solve the UK's economic problems."	3	3	3	3
"It is the government's responsibility to provide a job for everyone who wants one."	3	3	2.5	3
"The NHS primarily exists to look after those who can't afford private health care."	2	2	2	2
"The NHS will never have enough resources to completely satisfy all the demands made of it."	4	5	5	5
"The NHS needs to manage the resources that it has far more efficiently."	5	5	4	5
"Private health care providers offer greater choice to patients than the NHS does."	3	3	3	3
"Anyone who can afford it should take out private medical insurance."	3	3	3	3
"The poor are entitled to more help."	3	3	2	4
"It is essential that NHS services are targeted to the needs of the local population."	4	4	4	5
"There is one law for the rich and one for the poor."	3	3	3	3
"Private health care providers can assist the NHS in meeting patient's needs."	4	4	4	4
"Much of what the NHS does is unnecessary."	3	3	2.5	3
"People on higher incomes already pay more taxes, so are just as entitled to free treatment."	4	4	4	4
"Patients should receive public health care when needed and regardless of costs to the NHS."	4	4	4	4
"Big business benefits owners at the expense of workers."	3	3	3	3
"People who use the health service more often than average should pay more."	2	2	2	1
"Private health care should be abolished."	2	2	2	2
"The government should spend more on the NHS so that it can avoid having to make difficult decisions about who gets treatment and who doesn't."	4	4	4	4

1 = Strongly Disagree, 2 = Somewhat Disagree, 3 = Neutral, 4 = Somewhat Agree, 5 = Strongly Agree

The results of the factor analysis on the health care belief questions are presented in Table 4.4. The bigger the numbers in the table, the more important the question is in determining the value taken by the specific factor. Positive and negative figures show whether this figure is correlated with or against the factor. Bartlett’s test of sphericity (p = 0.000) suggests that factoring is appropriate, although the Kaiser-Meyer-Olkin measure (0.659) is only mediocre so that significant variation remains after factoring. 17 of the 20 questions are used by the four factors identified in the analysis, which we call “equity”, “market support”, “mixed” and “NHS entitlement”.

The “equity” factor is stronger for those agreeing to the statements suggesting that the rich benefit at the expense of the poor and that the government should give greater help to the poor at the expense of the rich. Those agreeing to these statements also tend to suggest that this should not necessarily extend to means-tested access or copayments on NHS treatment, since they also agree that "People on higher incomes already pay more taxes, so are just as entitled to free treatment”.

Table 4.4: Large coefficients (> 0.40) in factor analysis of belief questions

	EQUITY	MARKET SUPPORT	MIXED	NHS ENTITLEMENT
1. "Government should redistribute income from the better-off to those who are less well-off."	0.717			
10. "The poor are entitled to more help."	0.701			
12. "There is one law for the rich and one for the poor."	0.658			
15. "People on higher incomes already pay more taxes, so are just as entitled to free treatment."	0.478			
17. "Big business benefits owners at the expense of workers."	0.411			
9. "Anyone who can afford it should take out private medical insurance."		0.660		
8. "Private health care providers offer greater choice to patients than the NHS does."		0.630		
3. "Private enterprise is the best way to solve the UK's economic problems."		0.558		
18."People who use the health service more often than average should pay more."		0.453	- 0.505	
13. "Private health care providers can assist the NHS in meeting patient's needs."		0.425	0.474	
19. "Private health care should be abolished."		- 0.439	- 0.418	
6. "The NHS will never have enough resources to completely satisfy all the demands made of it."			0.540	
14. "Much of what the NHS does is unnecessary."			- 0.559	
16. "Patients should receive public health care when needed and regardless of costs to the NHS."				0.660
7. "The NHS needs to manage the resources that it has far more efficiently."				0.614
20. "The government should spend more on the NHS so that it can avoid having to make difficult decisions about who gets treatment and who doesn't."				0.476
11. "It is essential that NHS services are targeted to the needs of the local population."				0.453

"Market support" suggests general agreement with the principles of private enterprise and private health care – that health insurance should exist, that those with sufficient funds should use health insurance, that health insurance helps patients and the NHS, and that some copayment in the use of health care is desirable (so that those using the NHS more should contribute more). The "NHS entitlement" factor suggests strongly that whilst there is inefficiency in the NHS ("The NHS needs to manage the resources that it has far more efficiently"), there is also cause to increase the NHS budget to reduce the impact of scarce resources. The statement that "It is essential that NHS services are targeted to the needs of the local population" could be interpreted as an inefficiency or scarcity argument consistent with the above.

The "mixed" factor is the least effective explanatory of variation between responses. It argues against abolishing private health care and argues that such care can assist the NHS, but also argues against copayments and the argument that "Much of what the NHS does is unnecessary". The argument that "The NHS will never have enough resources to completely satisfy all the demands made of it" could be interpreted as either a justification of private provision or an argument in defence of the NHS. In either case, it is consistent with the idea that this factor provides support for a mix of public and private provision.

These four belief factors were regressed against background variables and dummy variables representing both study format and stage (IO, RI pre-discussion, RI post-discussion, RI pre-interview). F tests reveal a significant baseline difference for the equity factor between the IO and RI pre-discussion responses. In both cases, the IO response tends to display a higher value on this factor than the RI responses, even after adjusting for background variables. (In this case sex, age, education, and whether the individual has health insurance were all significant influences on the equity factor.) No significant effect was detected in equity between the questionnaire timings within the RI design. There are no other significant differences between the groups in the pre-interview and baseline comparisons on the other three factors.

Overall, the effect of different timings in the RI design is somewhat weak. The pre- and post-discussion coefficients do differ, but not significantly. For all four factors, the period of deliberation following the group discussion had the opposite effect to the group discussion. This suggests that the group process may have an impact but it seems unlikely to be a long-lasting or permanent one. This makes it less likely that the use of group discussions will lead to the elicitation of stable preferences.

4.2.2.3 Differences in willingness to prioritise

Table 4.5 presents the median values for the prioritisation questions. Across the groups, the median preferences suggest that it is considered more important to treat those who have waited longer, are in worse health, and have NHS-caused illnesses. In contrast, most people felt it was largely unimportant to give priority to those with lower incomes or in paid employment. The table shows that there were some small changes in the responses to the prioritisation questions but the general effects are unclear and the duration of their effect is uncertain.

Table 4.5: Prioritisation questions: median responses by observation set

	Median			
Where there are no other differences between people, it is more important for the NHS to treat	RI pre-discussion	RI post-discussion	RI pre-interview	IO pre-interview
(1)… those who have been waiting longer.	5	4.5	4	5
(2)… those who are in worse health.	5	4	4	5
(3)… those who have lower incomes.	1	2	2	2
(4)… those who are in paid employment.	2	2	2	2
(5)… those with children to look after.	3.5	3	4	3
(6)… those with elderly relatives to care for.	3	3	3	3
(7)… children and teenagers rather than adults.	3	3	3	3
(8)… younger adults (e.g. 25-40 year olds) rather than the elderly (e.g. 60-80 years olds).	2	3	3	2
(9)… those whose illnesses are caused by the NHS.	4	4	4	4
(10)… those whose illnesses are of genetic origin.	3	4	3	3

5-point Likert Scale: 1 = Not at all important, 3 = Neutral, 5 = Very important

The results of the factor analysis on the prioritisation questions are presented in Table 4.6. The factor analysis identified three factors using nine of the ten questions asked.

Again, the bigger numbers in the table, the more important the question is in determining the value taken by the specific factor. Bartlett's test of sphericity (p = 0.000) again suggests that factoring is appropriate, and more variation is explained by the factoring (Kaiser-Meyer-Olkin measure = 0.758) in this case.

If delay of treatment worsens health, then the two main factors in "prioritise clinically" lead to a concentration on the level of health when left untreated. The second factor "prioritise generally" suggests that there may be a tendency to support NHS prioritisation as a disparate list of people benefit here – those caring for children or elderly relatives, those with lower incomes, those in paid employment (which appears to conflict with the previous factor), and those with genetic illnesses. The final factor suggests that benefits to the young should be prioritised, as it includes both questions explicitly prioritising the young and the case in which the young are benefited by having healthier carers.

Table 4.6: Prioritisation questions: factor analysis and comparisons between observation sets within factor regressions

Where there are no other differences between people, it is more important for the NHS to treat	Prioritise clinically	Prioritise generally	Prioritise young
(1)… those who have been waiting longer.	0.713		
(2)… those who are in worse health.	0.814		
(6)… those with elderly relatives to care for.		0.787	
(3)… those who have lower incomes.		0.768	
(4)… those who are in paid employment.		0.744	
(10)… those whose illnesses are of genetic origin.		0.518	
(5)… those with children to look after.		0.707	0.464
(8)… younger adults (e.g. 25-40 year olds) rather than the elderly (e.g. 60-80 years olds).			0.830
(7)… children and teenagers rather than adults.			0.766
Significant background variables (in regression):	EQUITY, MIXED, Gender	EQUITY, MARKET SUPPORT	MARKET SUPPORT, NHS ENTITLEMENT, Race

All tests F tests following removal of variable(s) from the regressions predicting prioritisation factor values. Positive values indicate more support for the first item listed in each comparison

Regressions were conducted for individual factor values by study format and staging. After correcting for background and health and political beliefs (see Table 4.6 for significant variables), no significant differences were found between either IO versus RI pre-discussion responses or IO versus RI post-discussion responses. This suggests that the choice of design may make relatively little difference at an aggregate level.

We also test the effect of the discussion and deliberation factors within the RI design. "Prioritise clinically" receives a drop in support (-0.289, which is insignificant) following the initial discussion, and this effect is weaker (-0.130) following further deliberation. "Prioritise generally" receives significantly more support (+0.386, p = 0.037) following the initial discussion, but this effect is weakened by deliberation (to 0.256, which is insignificant), with no significant difference between the baseline and pre-interview support. "Prioritise young" follows a similar pattern, with an initial significant increase in support (+0.444) weakened by deliberation (to 0.245) and no significant difference between baseline and pre-interview figures.

4.3 CONCLUSION: CHOOSING THE DESIGN FOR PHASE 3

Table 4.7 presents a summary of our conclusions. A simple comparison of preferences from the main pairwise elicitation tasks suggests that there are differences between the IO and RI designs of the questionnaire. We have also found background differences between the samples, the RI sample tending to be healthier than the IO sample, and the IO group appearing to be more in line with the BHPS dataset. Such background differences could affect both the overall health and political beliefs of each sample, and their consequent willingness to prioritise some groups over others. This would also be expected to affect data from the preference elicitation task.

Table 4.7: Interpretation of Phase 2 Results

Question	Conclusion
Do differences exist in preference elicitation task data between the IO and RI formats?	Yes.
Are there differences in background characteristics across the designs?	Yes.
Over and above background effects, are there differences in the elicited health and political beliefs?	Yes
Over and above background effects and belief effects, are there differences in the willingness to prioritise?	Yes, but only within RI.

A factor analysis was run to reduce the 20 political and health belief questions to four factors in a pooled sample of the three RI and the one IO administration of the self-completion beliefs questionnaire (“equity”, “market support”, “mixed”, “NHS entitlement”). Individual level scores were then assessed across groups using a regression that accounts for background differences. We found that background affected the support to the individual factors. After correcting for background differences, there was weaker support for the “equity” factor in the RI group at baseline (versus IO) but no significant changes in this support following the discussion, or between the discussion and interview. As such, it seems to be that the difference observed initially is not related to the design.

Individual willingness to prioritise was also assessed using factor analysis, with the 10 prioritisation questions reduced to three general factors (“prioritise clinically”, “prioritise young”, “prioritise generally”). On regressions removing the effects of both background and health and political beliefs, we found no significant differences *between* IO and RI per se but did find differences *within* the RI designs. Overall, design appears to have had no significant effect on the willingness to prioritise when an interview takes place in each design.

Deliberation does have an effect on general prioritisation preferences, but this appears to be limited to within the RI design rather than between the designs. Here, the group discussion significantly increases the tendency to prioritise in general, and to prioritise the young in particular. By the time of an interview, these effects are no longer significant and hence are possibly less important in the choice of design (although 45-66% of this effect still remains at the time of interview). Given the small sample size of the RI framework and the lack of power implied by this, it may be unwise to automatically infer that deliberation is unimportant and particularly given that some of these pre-interview comparisons approach significance.

This smaller sample size makes the choice of design more complex. If the effects above are unimportant, then design has relatively little impact on the willingness to prioritise once we take account of background characteristics. Here, the differences we observe in the Thurstone values may be largely due to background characteristics and therefore best ignored in considerations of which design to select. Here, the additional group discussion and personal deliberation in the RI design arguably has little effect on preferences once an interview would take place.

At a practical level, the IO format managed a more representative mix of the population suggests that it may be simpler to obtain the spread of the population desired in the final sample. Recruitment in the RI was problematic: of the 116 people who agreed to participate in the RI design, for only 70 could a date be made with the interviewers, and only 57 of these attended a discussion group, and 56 answered the subsequent interview. The difficulty appears to lie in the logistics of the discussion group rather than a loss of interest in the deliberation period. The RI design is more expensive than the IO design, since it requires increased compensation to the respondents, increased interviewer costs, and the hire of venues. All three reasons recommend the use of the IO design.

The results obtained may also be stronger under the IO design. Since the numbers recruited to the IO design were increased in Phase 2 to allow comparisons to be made, the use of RI at Phase 3 would require operating at a decreased sample size in our final analysis, since we were unable to obtain the required sample at Phase 2. In contrast, the use of the IO design would take advantage of the oversampling at Phase 2. Use of the IO design would also allow additional information to be gathered within the project funding. We therefore use the IO design at Phase 3.

CHAPTER 5: GENERATING EQUITY WEIGHTS

Phase 3 of this project was the main study designed to parameterise the social welfare function. It is based on the preference elicitation task in the interview-only (IO) design of Phase 2, and so we can include the results from Phase 2 in the analysis reported here. Moreover, resources were available for a smaller scale additional study, which looked further at the issues of timing and severity, and this chapter reports also reports on those results.

5.1 THE MAIN STUDY

5.1.1 Methods

Please refer back to Table 4.1 for the question used in the main study.

5.1.1.1 Recruitment and respondents

Our sample size was determined by being able to represent the preferences of a sample of the UK general population and by being able to show differences by important population sub-groups. The central limit theorem can be used to assume an approximately normal distribution for a sum or mean of preferences (for instance, the number of times a particular study state is preferred across four choices). The general rule of thumb in such cases is that at least 30 observations must be obtained for each subgroup for the central limit theorem to be used. (A limit of 30 observations per subgroup is also suggested by Louviere et al (2000) for the analysis of DCEs.)

We consider population subgroups divided according to:

1) Gender
2) Schooling: whether the respondent experienced higher or further education, or schooling up to or including secondary school
3) Age: those aged 16-40, 40-59 and 60+.

If we aimed to obtain at least 30 respondents in each of these 12 (2×2×3) combinations of background characteristics, we require a *minimum* sample size of 360. It was felt that the actual sample size necessary to obtain achieve this would therefore be larger than 360. It was difficult to know how much allowance needed to be made here, and an additional 180 (50%) interviews were obtained over and above the 360. As not all respondents provided full data, we aimed to recruit a total of 580 individuals by the end of Phase 3.

Eight interviewers were used in Phase 3, seven of which had taken part on Phase 2. There were a total of nine interviewers across the two phases. As before, all households were sent a participation information sheet and covering letter detailing the study before any contact by an interviewer. Not all those receiving a letter were contacted by an interviewer. Respondents agreeing to participate were paid £5 for doing so.

The background questions we used were designed for compatibility with the BHPS, so that comparisons with the BHPS provide the simplest measures of representativeness (as in Chapter 4). The BHPS may not be entirely representative due to demographic changes over time, so we compare the proportions in each of our population subgroups to equivalents drawn from the 2001 Census. Key statistics from the Office of National Statistics provide a crosstab breakdown of sex and age by the highest level of qualification for those aged 16-74[1]. This breakdown is not available for those aged 75 years and above, so we assume a similar educational breakdown those aged 60-74 years of age. Population subgroups are estimated using a crosstab breakdown of sex and age for those aged over 75 years.[2] Our estimated population subgroup proportions are given in Table 5.1 and we compare our sample to this table below.

Table 5.1: Estimated subgroup proportions in 2001 Census.

Gender	Education	Age		
		16-40	40-59	60+
Male	School only	15%	12%	9%
	HE/FE	5%	4%	2%
Female	School only	16%	13%	13%
	HE/FE	5%	4%	2%

5.1.1.2 Analysis

5.1.1.2.1 Aggregate estimates

The homotheticity property of the CES function also allows us to make predictions about societal preferences regarding different distributions of health. In Question 1, the groups are identical except in the health they receive and so α =0.50 is assumed. Given this assumption, our analysis for Question 1 (Choice Sets 1.1 to 1.4) finds the r values that correspond to a SWF that assumes that the standard QALY model holds. As Choice Sets 1.3 and 1.4 were derived by halving the total number of life years enjoyed in Choice Sets 1.1 and 1.2 in all states, then it also halves the number of QALYs received. If the number of QALYs describes how society judges individual lifetime health, then the CES function predicts that the same amount of inequality aversion should be exhibited in both cases. If this does not hold then either the QALY does not measure individual preferences or homotheticity does not hold.

When analysing Sets 1.1 to 1.4, we assess the consistency of individual choices against the predictions of the CES-SWF. We do this by comparing responses across questions to look for differences in inequality aversion in similar-sized questions (1.1 vs. 1.2; 1.3 vs. 1.4), and when homotheticity is tested (1.1 vs. 1.3; 1.2 vs. 1.4). Choice Sets 2.1 to 2.4 are used to construct both an alternative to the standard QALY when considering lifetime health from a societal perspective and to estimate the degree of inequality aversion exhibited in the data. These choices are analysed in terms of "Adult Healthy Year Equivalents" or AHYEs, v_1 and v_2, which like the QALY combine information regarding both health (quality of life) and the timing of health. In Question 2, we use $\alpha = 0.5$ as Groups 1 and 2 are identical in all respects except the health they receive.

---

[1] Table S105, http://www.statistics.gov.uk/StatBase/Expodata/Spreadsheets/D7532.xls

[2] Using the age/sex gender breakdown in Table S001 http://www.statistics.gov.uk/StatBase/Expodata/Spreadsheets/D7547.xls.

The timing variable takes two values, and distinguishes between the health experienced prior to 18 years of age and all health experienced at and above this age. Health takes three levels, being dead, 25% and 100% health. These judgements may be consistent with conventional unweighted QALYs but they may give different weights to ill health at different ages than suggested by the QALY model. The AHYE reflects these judgements and values a profile of health using the number of years in full health as an adult that would be equivalent to it. Where the conventional unweighted QALY model adequately describes societal health judgements, the two concepts coincide.

Given the study and Thurstone-equivalent states for Choice Sets 2.1 to 2.4, we can estimate a basic form for the SWF (using AHYEs) in the case where the groups are identical in non-health respects ($\alpha = 0.5$). Earlier choice sets from Question 1 are ignored here because they do not consider periods of ill-health, and there is a concern that this may affect the responses given. (Later choice sets from Questions 3 onwards are ignored at this stage because α will not typically equal 0.5.) Of interest here is whether, and how, these estimates differ from the normal assumptions of cost-effectiveness analysis:

- Do overall preferences exhibit inequality aversion – that is, is the inequality parameter (r), typically greater than negative one?
- Is there extra weight placed on the health of children versus those of adults – that is, does $FHC = FHA = 1$?
- Is there a premium (discount) placed on the value of 25% health over the 25% that cost-effectiveness analysis assumes?

Within the general form of the SWF, the parameter α is used to capture the degree to which groups are treated differently for non-health reasons. In order to assess the effect of condition cause/responsibility, we compare cases where this information is provided, with otherwise identical cases where it is not. The states in Choice Set 2.1 also appear as 3.1, 4.1, 5.1 and 5.2. The states in Choice Set 2.3 appear as 3.2 and 4.2. In each case, the condition cause choice sets in questions 3 and 4 are compared to the baseline choice set in question 2.

The trade-offs where health differences do not exist between the groups – i.e. where only α may differ – are found by again comparing the results from Choice Sets 2.1 and 2.3 with those of Choice Sets 3.1-5.2. For Choice Sets 2.1 and 2.3, we solve to find the α providing identical social welfare values between the study and equivalent states in each case. This provides a baseline figure (α_0) correcting for any residual error in the question – which is likely to exist because the main estimates of r, FHC and SHA are based reducing error across Choice Sets 2.1-2.4 as a whole. For these baseline cases, the marginal rate of substitution in the absence of health differences equals ($\alpha_0/(1-\alpha_0)$).

We also solve for the individual parameter values (α_1) for Choice Sets 3.1-4.2, with corresponding marginal rates of substitution ($\alpha_1/(1-\alpha_1)$). The effect of the condition cause label (in the absence of health differences) is the ratio of the two marginal rates of substitution. Similarly, by comparing Choice Sets 3.1 and 5.1, and 4.1 and 5.2, we can consider the impact of labels versus the more abstract descriptions. Condition rarity is considered in Question 6. As with the condition cause questions above, we can define Choice Set 6.1 – where both groups suffer from an equally rare condition – as the baseline and Choice Set 6.2 – where the same numbers are used but one condition is extremely rare.

Bootstrapping is used to assess the uncertainty in the parameter estimates. Microsoft Excel was used to resample the data, with a random number generator to identify rows within the database containing complete data on Choice Sets 1.1 to 6.2 in the preference elicitation task. By sampling the same number of rows as we have individuals with complete data, we can find a new dataset that reflects a similar level of heterogeneity as the original dataset. By repeating the analysis (finding equivalent states and parameter values) multiple times, we can estimate the uncertainty in each parameter value. For the main parameter estimates, a sample of 5000 observations was used, as this was considered likely to allow convergence when estimating of parameter uncertainty. Convergence was assessed by examining how quickly parameter estimates of standard deviation reached its final value. This was then used to estimate the necessary number of iterations in subsequent analyses.

5.1.1.2.2 Effect of background characteristics and sensitivity analysis

We also consider whether the SWF parameters differ according to background characteristics in the main study. In each of the 12 respondent characteristic group, values for the parameters *r, FHC* and *SHA* (with $SHY = FHY \times SHA$) are found using preferences from Choice Sets 2.1-2.4. Uncertainty is again computed using bootstrapping, and comparisons are made across:

- Each of the 12 respondent characteristic subgroups differentiated
- Those judging themselves to be either in “excellent” or “good” health (health judged as) versus those who judged themselves to be in poorer health (“fair”, “poor” or “very poor”). (Median response = good).
- Those who judge themselves to be disabled, versus those who judged themselves not disabled
- Those who had, and did not have, private health insurance.
- Those who scores higher than average on each of the four health and political belief factors found in Phase 2 analysis, versus those who do not.

If differences are found by background characteristics, then the main analysis results may depend on the makeup of the sample itself. We therefore construct a virtual sample that is broadly representative of the general public where background differences affect preferences. In this sample, we require that the twelve background groups (gender × education × age) should be selected in the proportions they appear in the general public. This can be done using bootstrapping methods to provide both a central estimate for parameter values and uncertainty. We therefore re-run our analyses by resampling to our original sample size but require that each population subgroup provides the “correct” number of respondents when split by age, sex and education.

In the main estimates all individuals are included in the dataset so long as they give answers to all questions in Choice Sets 1.1-6.2. We consider the effects of introducing rationality criteria for inclusion in the analysis using two concepts. For an individual, we can say that an individual is “rank-consistent” where their preference for the study state (x_5) is 1) non-decreasing as we move from x_1 to x_4 and 2) indifferent to at most one reference state. Where the former restriction is violated, the responses are labelled “non-monotonic”, and those preferences violating the second criteria are labelled as “indifference violations”. We can also identify those respondents (on Choice Sets 1.1-1.4) whose preferences exhibit inequality aversion or inequality neutrality. These concepts are combined to find the number of choice sets for each individual that are consistent with our basic SWF. Data quality may be improved if those frequently choosing in a rank inconsistent way or preferring inequality are removed from the dataset. The effect of varying the threshold in relation to these conditions is explored.

Within the main analysis of the project, we assume that all health below the age of 18 is “childhood” health, and all health after the age of 18 is “adult” health. Whilst the split between adult and childhood health is necessary within the analysis, the precise cut-off does not necessarily fall at 18. Indeed, the main study includes questions where all ill-health during childhood is endured from birth up to a maximum of eight years old. We therefore explore the effect of modifying this cut-off to 10 and 20 years of age.

When using Thurstone scores, pairwise choices where one point is preferred all the time present a problem, since here the scores become indeterminate. The typical response to this issue is either to omit such data or restrict proportions to fall within a permitted range. Guildford (1954) recommends using proportions in the range 2.3%-97.7%, whilst Edwards (1957) uses thresholds at 1% or 2% (plus 98% or 99%). We use a 2%-98% range within our main analysis. Since this cut-off is somewhat arbitrary, we also analyse the alternative cases where 1% and 5% threshold limits are set (upper limits: 99%, 95%). This will typically only impact on the inferred preferences between the reference states in any choice set.

5.1.2 Results

5.1.2.1 Recruitment and respondents

The breakdown of our sample by population subgroup is given in Table 5.2. Overall, we obtained complete data on all pairwise questions from 559 of 582 respondents of which 225 (of 232) were obtained in Phase 2 and 334 (of 350) were obtained in Phase 3. Where Phase 2 provided few respondents in a particular group the interviewers were instructed to target these respondents wherever there was a choice of respondents in a particular household. As a result, we successfully recruited over 30 individuals to each of the population subgroups (with a minimum of 38 for more educated, elderly males)

Table 5.3 reports the background of the sample against data from the 2001 Census (or closest equivalent). In general, the sample is roughly representative in terms of age and gender (although our sample includes more 60-69 year olds). As in Phase 2, we slightly under-represent non-white ethnicities and those with disability/chronic illness, and over-represent the retired and those with higher or further education. Note, however, that the disability comparison is slightly different (“Do you consider yourself to be a disabled person” versus limiting long term illness. In any event, given that we correct for non-representative preferences in the sensitivity analysis (on age, sex and education), any non-representativeness is not a major concern.

Table 5.2: Sample characteristics (complete data only) against 2001 Census data

Gender	Education	Age	Phase 2	Phase 3	Total		2001 Census
					N	%	
Male	School Only	<40	4	40	44	8%	15%
		40-59	15	23	38	7%	12%
		60+	13	30	43	8%	9%
	HE/FE	<40	13	32	45	8%	5%
		40-59	22	28	50	9%	4%
		60+	20	17	37	7%	2%
Female	School Only	<40	11	34	45	8%	16%
		40-59	14	23	37	7%	13%
		60+	27	29	56	10%	13%
	HE/FE	<40	32	36	68	12%	5%
		40-59	31	25	56	10%	4%
		60+	23	16	39	7%	2%
Total			225	334	558[*]	100%	100%

*One respondent did not have complete subgroup data.

Table 5.3: Background of the sample

Sample size		Social QALY sample (%)	2001 Census (%)
Gender:	Female	55	52
Age:	18-29	20[a]	19
	30-39	16	20
	40-49	17	17
	50-59	15	16
	60-69	17	12
	70+	15	15
Ethnicity	White	95	92[b]
Employment status:	Self-employed	7	8
	Other Employed	39	52
	Retired	29	14
Education:	School only	47	78b
	HE/FE	53	22[b]
House ownership:	Owned/mortgage	71	71[c]
Disabled?	Yes	14	18[d]

[s] Includes 14 aged below 18.

[b] Ages 16-74 only

[c] 2000 data. Office of National Statistics.

[d] Limiting long-term illness .

Overall, nine interviewers were used to obtain the 559 interviews with complete data. (The rate of incomplete interviews by interviewer ranged from 0-10%, against an average of 4%). Our data was obtained from 17 areas, with numbers in each varying from 10 (Sheffield) to 40 (both Oldham and Merthyr Tydfil). Overall, 104 respondents were obtained from NW England, 34 from NE England, 206 from Yorkshire and the Humberside, 101 from the Midlands, 65 from the South of England

and 72 from Wales. Whilst we selected a range of geographical locations, our emphasis was to obtain a sufficient sample across the range of gender, age, and education.

5.1.2.2 Overall ranking results

Table 5.4 shows the proportion of the 559 respondents with complete data that prefer the study state (x_5) to the reference states (x_1 to x_4). The proportion preferring the reference state can be found as one minus these figures (where an individual is indifferent between the study and reference states the preference is split equally between the reference and study states). These preferences are used to create the Thurstone scores.

Since $x_1 \succ x_2 \succ x_3 \succ x_4$ holds by assumption, we would expect the study state (x_5) to be preferred to the best reference state (x_1) less often than it is preferred to the worst state (x_4). In other words, by reading along the rows from left to right, we expect the percentages to increase. At an aggregate level, the average preference has $x_1 \succ x_2 \succ x_5 \succ x_3 \succ x_4$ in all but one case (Choice Set 5.1).

Table 5.4: Proportions preferring study state by reference state and choice set

	vs. Reference State			
Choice Set	x_1 (best)	x_2	x_3	x_4 (worst)
1.1	13%	37%	77%	77%
1.2	18%	33%	64%	78%
1.3	11%	39%	84%	86%
1.4	24%	40%	73%	83%
2.1	37%	38%	42%	48%
2.2	43%	46%	47%	52%
2.3	36%	38%	43%	47%
2.4	39%	43%	46%	50%
3.1	42%	48%	53%	57%
3.2	45%	49%	53%	56%
4.1	35%	48%	49%	56%
4.2	42%	47%	49%	55%
5.1	50%	54%	59%	58%
5.2	35%	44%	48%	55%
6.1	45%	47%	52%	57%
6.2	46%	49%	52%	60%

Aggregate level data.
Preference for study state expected to increase across columns.

In any given choice, a rank-consistent set of responses requires that it is possible to derive a clear and consistent rank ordering between the study state and reference states. Table 5.5 presents the number of rank-consistent responses by question and choice set.

A majority of individuals answer in a rank-consistent fashion on all choice sets. The numbers of rank-consistent responses are typically higher after the first question in each choice set, which is consistent with respondents learning the task. However, as the questionnaire proceeds, the number of individuals choosing in a rank-consistent fashion tends to fall. This may be due to the complexity of the questions – this increases at Question 2 (introduction of poor health, fewer monotonic responses), Q3 (introduction of condition cause) and Q5 (introduction of labelling). Question 6 is largely stand-alone and is similar in style – and rank consistency – to Question 1. Approximately one-third of individual responses are not consistent with a rank order. Whilst this figure may sound high, it is worth noting that these orders are the result of four individual pairwise choices and errors in *any* of these choices can prevent a consistent rank-ordering. The chance of making an error on any particular choice is obviously much lower.

Table 5.5: Rank consistency by question and choice set

N	x.1	x.2	x.3	x.4	ALL
All Q1	79%	86%	85%	84%	83%
All Q2	57%	65%	73%	74%	67%
All Q3	56%	72%	▨	▨	64%
All Q4	63%	74%	▨	▨	68%
All Q5	57%	64%	▨	▨	60%
All Q6	85%	87%	▨	▨	86%

Shaded area indicates that the question does not have a corresponding choice.

The responses to Choice Sets 1.1 and 1.4 were coded to form variables indicating how often the study state was preferred to the four reference states in a rank-consistent fashion. Rank consistent responses were given numerical values ranging from 4 (study state preferred to all four reference states) to 0 (study state preferred to none of the reference states). Cases of indifference were valued according to the number of strictly preferred states plus a half. For example, the preference ordering $x_1 \succ x_2 \sim x_5 \succ x_3 \succ x_4$ would be coded as 2 + 0.5 = 2.5. Non-rank consistent responses were coded as “NR”.

If the QALY model holds as representation of preferences over health profiles and if homotheticity holds, then the responses to Choice Sets 1.1 and 1.3 should be similar because Choice Set 1.3 halves the figures used in Choice Set 1.1. In the same way, the responses to 1.2 and 1.4 should also be similar. Chi-squared tests show a strong and significant difference between the results of both pairs of choice sets ($p < 0.001$). Of the individuals giving rank-consistent preferences to Choice Sets 1.1 and 1.3, 51% (n = 201/393) choose in exactly the same way across the two choices, whilst 93% (n = 367/393) changed no more than one choice. Of the individuals giving rank-consistent preference to Choice Sets 1.2 and 1.4, 47% (n=204/431) gave exactly the same response to both questions and 85% (n=367/431) changed no more than one choice.

The study states in Choice Sets 1.1 and 1.3 involves the same total number of QALYs, and less inequality, than in Choice Sets 1.2 and 1.4 respectively. Since the same reference states are used in each choice set, we would expect that the reference state in Choice Set 1.1 (1.3) to be at least as good as the reference state in Choice Set 1.2 (1.4). Of those with rank-consistent preferences across each pair of responses, 71% (n=287/404) display such preferences in a comparison of Choice Sets 1.1 and 1.2. For Choice Sets 1.3 vs. 1.4, the corresponding figure is 73% (309/425).

Table 5.6 presents the scaled Thurstone scores for the study and reference states by choice set. These figures show how social welfare changes as we move from the best reference state to the worst (from x_1 to x_4), and where the study state (x_5) lies in relation to the reference states. The figures are anchored so that W(x_2) and W(x_3) equal 1 and 0, respectively. In aggregate, all the values for x_5 are placed between x_2 and x_3, and hence obey the aggregate preference ordering $x_1 \succ x_2 \succ x_5 \succ x_3 \succ x_4$. The value for W($x_5$) shows where, in preference terms, the study state is located between x_3 and x_2.

Table 5.6: Scaled Thurstone scores by state and choice set

	Reference states				Study state
	W(x_1) (best)	W(x_2)	W(x_3)	W(x_4) (worst)	W(x_5)
1.1	1.945	1.000	0.000	-0.789	0.540
1.2	1.924	1.000	0.000	-0.916	0.450
1.3	1.941	1.000	0.000	-0.779	0.679
1.4	1.924	1.000	0.000	-0.897	0.654
2.1	1.987	1.000	0.000	-1.016	0.230
2.2	2.016	1.000	0.000	-1.021	0.403
2.3	1.984	1.000	0.000	-0.992	0.229
2.4	2.001	1.000	0.000	-1.006	0.336
3.1	2.016	1.000	0.000	-1.004	0.503
3.2	1.999	1.000	0.000	-0.995	0.520
4.1	2.079	1.000	0.000	-1.044	0.417
4.2	2.017	1.000	0.000	-1.021	0.449
5.1	1.994	1.000	0.000	-0.968	0.663
5.2	2.034	1.000	0.000	-1.022	0.363
6.1	1.976	1.000	0.000	-0.993	0.505
6.2	2.001	1.000	0.000	-1.027	0.553

Aggregate level data.
Scaled Thurstone scores represent a cardinal measure of social welfare.
Scores for CS1.1-1.4 are expected to lie between 0 and 1 due to inequality aversion.
Scores for other choice sets can take any value.

5.1.2.3 Inequality aversion

The Thurstone scores all place the study states between x_2 and x_3 in preference terms, and this is consistent with the CES-SWF at an aggregate level. The scaled Thurstone score for the study state W(x_5) can also be used to define an “equivalent” state:

$$x_6 = \mathrm{W}(x_5)\,x_2 + (1-\mathrm{W}(x_5))\,x_3$$

The Thurstone equivalent state lies along the same line as x_1 through to x_4. Table 5.7 gives the study states and equivalent states that would be valued equally for each choice set. Within Choice Sets 1.1 to 1.4, the aggregate preferences suggest a trade-off between total health and reducing inequalities. In Choice Set 1.1, for example, respondents are willing to sacrifice 5.84 QALYs (70.00 – 64.16 QALYs) to the better off group in order to obtain 2.16 QALYs (58.16 – 56.00 QALYs) for the worse off group. This suggests an implicit marginal rate of substitution between the health of the worst off to the best off of 2.7; that is, the health of the worst is worth 2.7 that of the health of the best off. For the other choice sets, this figure varies from 1.42 (Choice Set 1.4) to 1.94 (Choice Set 1.3).

Table 5.8 gives the inequality aversion parameters, *r*, and measures of uncertainty within this for Choice Sets 1.1 to 1.4 under the assumption that the QALY describes the way that society judges lifetime health. For example, the trade-off for Choice Set 1.1 suggests an inequality aversion parameter of $r = 5.24$. The central estimates for the *r* parameters appear to differ across the four choice sets. Consider a standard case where Group 1 has a life expectancy of 70 years in full health and Group 2 has a life expectancy of 60 years in full health. The inequality aversion estimate from Choice Set 1 would suggest that this is equivalent to case where both groups live 63.83 years in full health. Here, Group 1 loses 6.17 years and Group 2 gains 3.83 years, suggesting that across these improvements Group 2's health is worth 61% more than Group 1's health (6.17/3.83-1). For Choice Sets, 1.2, 1.3, and 1.4, the comparable figures are 21%, 38% and 13%, respectively.

Table 5.7: Question 1 Choice States and Equivalents

Choice Set & States		Group 1 Health	Group 2 Health	Trade-offs
1.1	Study State	70 years in 100% health	56 years in 100% health	2.70 QALYs (Group 1) per QALY (Group 2)
	Equivalent	64.16 years in 100% health	58.16 years in 100% health	
	Difference	- 5.84 QALYs	+ 2.16 QALYs	
1.2	Study State	74 years in 100% health	52 years in 100% health	3.64 QALYs (Group 1) per QALY (Group 2)
	Equivalent	63.80 years in 100% health	57.80 years in 100% health	
	Difference	- 10.2 QALYs	+ 2.80 QALYs	
1.3	Study State	35 years in 100% health	28 years in 100% health	1.94 QALYs (Group 1) per QALY (Group 2)
	Equivalent	32.36 years in 100% health	29.36 years in 100% health	
	Difference	- 2.64 QALYs	+ 1.36 QALYs	
1.4	Study State	37 years in 100% health	26 years in 100% health	2.03 QALYs (Group 1) per QALY (Group 2)
	Equivalent	32.31 years in 100% health	29.31 years in 100% health	
	Difference	-4.69 QALYs	+ 2.31 QALYs	

Results from samples of 5000 bootstrapped observations.
Study states defined by choice sets.
Equivalent states derived at an aggregate level using Thurstone scores.
Lifetime health judgements are assumed to equal the number of QALYs.
Differences derived in QALY terms.

Table 5.8: Inequality parameter (*r*) estimates by choice set

Choice Set	Mean	Minimum	Maximum	Std Dev	95% CI
1.1	5.24	3.64	6.96	0.50	(4.266, 6.216)
1.2	1.51	0.78	2.19	0.19	(1.139, 1.871)
1.3	3.16	1.68	5.01	0.45	(2.289, 4.041)
1.4	0.55	-0.05	6.96	0.18	(0.207, 0.895)

Estimates based on 5000 bootstrapped observations.
Parameters derived using given study and equivalent states.
CES function is assumed to use lifetime health judgements equivalent to QALYs.

However, the differences found between Choice Sets 1.1 and 1.3, and between Choice Sets 1.2 and 1.4 do appear to suggest a general violation of the CES-SWF in the case considered here, so that QALYs and AHYEs are expected to differ. We therefore relax the assumption that the QALY model is used by individuals and in such cases we would not expect these pairs of choice sets to produce the same values for *r*, since halving QALYs is unlikely to halve the societal value of that health. As an illustration, suppose that childhood health is valued ten times as much as adult health. Living for 30 years (= 18+12) in full health provides 18×10+12 = 192 AHYEs. Halving this, 15 years in full health provides 15×10 = 150 AHYEs. Where timing affects preferences, we cannot test homotheticity directly.

Table 5.9 shows both the study states and their social welfare equivalents in Choice Sets 2.1 to 2.4. Within each of the equivalent states, ill-health always occurs at the end of life so, for instance, the equivalent state in Choice Set 2.1 involves 59.23 years of full health, followed by 4.92 years in severe health. The trade-offs defined in these states are complex, as they involve periods in ill-health, periods in good health, health as children and health as adults. By changing the parameters in the SWF, we seek to find the solution that gives the least difference in social welfare across the pairs of states.

Table 5.9: Question 2 study states and equivalents

Choice Set & States		Group 1 Health	Group 2 Health
2.1	Study State	66 years in 100% health 8 years in 25% health	50 years in 100% health 16 years in 25% health
	Equivalent	59.23 years in 100% health 4.92 years in 25% health	54.46 years in 100% health 8 years in 25% health
2.2	Study State	66 years in 100% health 16 years in 25% health	4 years in 25% health 54 years in 100% health 4 years in 25% health
	Equivalent	59.40 years in 100% health 5.61 years in 25% health	54.81 years in 100% health 8 years in 25% health
2.3	Study State	72 years in 100% health 16 years in 25% health	48 years in 100% health 16 years in 25% health
	Equivalent	59.23 years in 100% health 4.92 years in 25% health	54.46 years in 100% health 8 years in 25% health
2.4	Study State	8 years in 25% health 72 years in 100% health 8 years in 25% health	48 years in 100% health 16 years in 25% health
	Equivalent	59.34 years in 100% health 5.34 years in 25% health	54.67 years in 100% health 8 years in 25% health

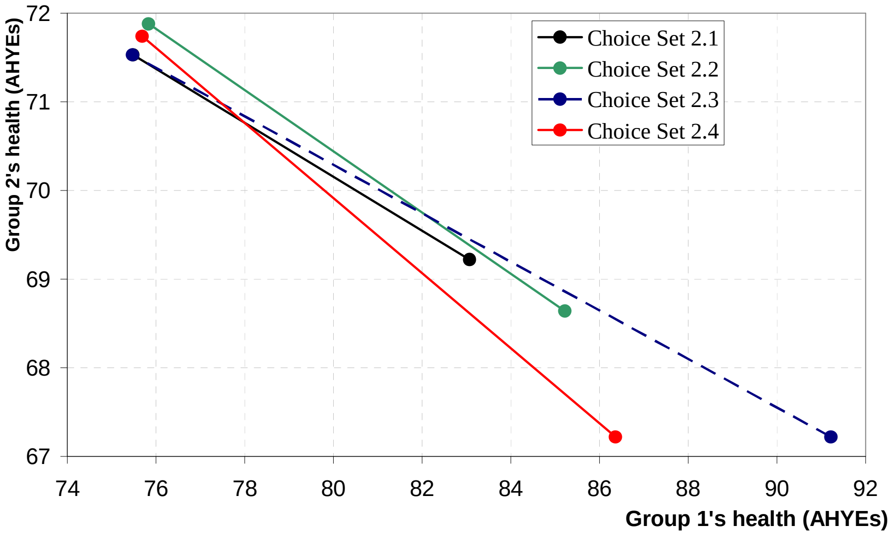
Results from samples of 5000 bootstrapped observations.
Study states defined by choice sets.
Equivalent states derived at an aggregate level using Thurstone scores.

Figure 5.1 plots the study and equivalent states together and it is clear that straight lines between these points do intersect (the two steeper lines (2.2, 2.4) are those where childhood illness may occur.) However, we can still note that there is a clear trade-off between average health and reducing inequalities in all cases, since the slope of all curves is less than one.

The mean parameters, their estimated standard deviations, and 95% CI are presented in Table 5.10, along with the assumptions of a standard QALY-based CEA. The standard CEA assumes no inequality aversion in its objective function ($r = -1$), whilst the SWF found here has significantly higher inequality aversion, with an inequality aversion significantly above 5. Consider a standard case where Group 1 has a life expectancy of 70 years in full health and Group 2 has a life expectancy of 60 years in full health; in terms of social welfare this is equivalent to a case in which both receive 64.76 years in full health.

Formally, this is analysed in terms of AHYEs where Groups 1 and 2 receive 84.9 (1.828*18 + 1.000*52) and 74.9 (1.828*18 + 1.000*42) AHYEs. This is equivalent to a case in which both receive 79.66 AHYEs (1.828*18 + 1.000*46.76). Group 1 loses 6.12 years and Group 2 gains 3.88 years of full health as an adult; over this change,

Figure 5.1: Choice Set 2 study state and equivalent points: AHYEs



Group 2’s health is valued 57% more highly than the health of Group 1. (Since both changes concern adult full health years the same trade-off is made regardless of whether we consider QALYs or AHYEs.)

Table 5.10: SWF parameters; standard CEA assumptions and study estimates

Choice Set	Label	CEA/QALY	Study Estimates	Standard Deviation	95% CI
Inequality aversion parameter	r	-1.00	6.32	0.29	(5.76, 6.88)
Lifetime health judgements					
Value of 100% health as an adult	FHA	1.000	1.000	-	-
Value of 100% health as a child	FHC	1.000	1.828	0.031	(1.768, 1.888)
Value of 25% health as an adult	SHA	0.250	0.268	0.012	(0.244, 0.292)
Value of 25% health to a child	SHC	0.250	0.490	0.027	(0.439, 0.542)

Estimates based on 5000 bootstrapped observations
Parameters derived using given study and equivalent states for CS2.1-2.4.
CES function is assumed to use lifetime health judgements weighted using FHA, FHC, SHA, SHC parameters (AHYEs).

The QALY values 25% health for an adult as worth 0.250 times as much as a full year in 100% to an adult: at 0.268, this weight is not significantly different in the AHYE. The QALY also values 25% health for a child as worth 0.250 times as much as a full year in 100% to an adult. In contrast, our results suggest that 25% health as a child is worth 0.490 AHYEs, which is significantly more than its QALY weight. Overall, the AHYEs gives 96% more weight to the first 25% health for children relative to that given by the QALY.

5.1.2.4 Condition cause/responsibility

Table 5.11 provides the study states and equivalent states up to Choice Set 4.2, along with summary information regarding the cause of illness. Each pair of study and equivalent states suggests a trade-off between the health of one group and the health of the other. For Choice Set 2.1, which is neutral with respect to the allocation of responsibilities and thus will be used as the baseline, Group 1 loses 6.77 years in 100% health and 3.08 years in 25% health, whilst Group 2 gains 4.46 years in 100% health whilst losing 8 years in 25% health. In terms of societal judgements of health, Group 1 loses 7.60 AHYEs (6.77 + 0.268×3.08) and Group 2 gains 2.31 AHYEs (4.46 – 0.268×8).

In Choice Set 3.1, Group 1 has an NHS-caused disease and Group 2 has a disease due to non-NHS causes that are partially due to the patient's lifestyle. Here, Group 1 loses 7.03 AHYEs and Group 2 gains 2.86 AHYEs. Compared to the neutral Choice Set 2.1, here in Choice Set 3.1 where condition cause/responsibility is included, the public is less willing to sacrifice health to Group 1 in order to gain more health for Group 2; a AHYE to Group 2 is worth 3.28 AHYEs to Group 1 in Choice Set 2.1, and 2.46 AHYEs to Group 1 in Choice Set 3.1. The difference between these figures is interpreted as being due to the condition cause/responsibility, with a higher value is placed on NHS-caused diseases than ones in which the patient is partially responsible.

Table 5.11: Equivalent states by condition cause/responsibility

Choice Set(s) and States	Group 1 Health	Group 2 Health	Trade-offs
2.1, 3.1, 4.1 Study state	66 years in 100% health 8 years in 25% health	50 years in 100% health 16 years in 25% health	
2.1 Equivalent state	(Baseline) 59.23 years in 100% health 4.92 years in 25% health	(Baseline) 54.46 years in 100% health 8 years in 25% health	3.28 AHYEs (Group 1) per AHYE (Group 2)
	-7.60 AHYEs	+ 2.26 AHYEs	
3.1 Equivalent state	(NHS causes) 59.50 years in 100% health 6.01 years in 25% health	(Non-NHS, patient) 55.01 years in 100% health 8 years in 25% health	2.46 AHYEs (NHS causes) per AHYE (Non-NHS, patient)
	-7.03 AHYEs	+ 2.86 AHYEs	
4.1 Equivalent state	(NHS causes) 59.42 years in 100% health 5.67 years in 25% health	(Non-NHS, non-patient) 54.83 years in 100% health 8 years in 25% health	2.68 AHYEs (NHS causes) per AHYE (Non-NHS, non-patient)
	-7.21 AHYEs	+ 2.69 AHYEs	
2.3, 3.2, 4.2 Study state	72 years in 100% health 16 years in 25% health	48 years in 100% health 16 years in 25% health	
2.3 Equivalent state	(Baseline) 59.23 years in 100% health 4.92 years in 25% health	(Baseline) 54.46 years in 100% health 8 years in 25% health	3.65 AHYEs (Group 1) Per AHYE (Group 2)
	-15.74 AHYEs	+ 4.31 AHYEs	
3.2 Equivalent state	(NHS causes) 59.52 years in 100% health 6.08 years in 25% health	(Non-NHS, partial patient) 55.04 years in 100% health 8 years in 25% health	3.09 AHYEs (NHS causes) per AHYE (Non-NHS, patient)
	-15.14 AHYEs	+ 4.90 AHYEs	
4.2 Equivalent state	(NHS causes) 59.45 years in 100% health	(Non-NHS, non-patient) 54.90 years in 100% health	3.22 AHYEs (NHS causes)

	5.80 years in 25% health	8 years in 25% health	per AHYE (Non-NHS, non-patient)
	-15.29 AHYEs	+ 4.75 AHYEs	

Estimates based on 5000 bootstrapped observations.
Study states defined by the choice sets.
Equivalent states derived at an aggregate level using Thurstone scores.
Trade-offs assume main SWF in AHYEs terms (Table 5.10).

In Choice Set 4.1, Group 1 has an NHS-caused disease and Group 2 has a disease due to non-NHS causes that are not due to the patient’s lifestyle. Here, Group 1 loses 6.58 years in 100% health and 2.33 years in 25% health (7.21 AHYEs), whilst Group 2 gains 4.83 years in 100% health and loses 8 years in 25% health (a gain of 2.69 AHYEs). Compared to the neutral Choice Set 2.1, here in Choice Set 4.1 where condition cause/responsibility is included, the public is again less willing to sacrifice health to Group 1 in order to gain more health for Group 2; a AHYE to Group 2 is worth 3.28 AHYEs to Group 1 in Choice Set 2.1, and 2.68 AHYEs to Group 1 in Choice Set 4.1. This suggests that more value is placed on NHS-caused diseases than ones in which illness is not due to non-NHS, non-patient causes.

Taken together, these two findings suggest that NHS-caused diseases may attract a higher value than the other two categories, with more value placed on diseases that are not due to patient lifestyle than those that are. Comparing Choice Sets 2.3 (no condition causes), 3.2, and 4.2 suggests a similar pattern.

Table 5.12 shows the significance of these weights. Relative to the case of non-NHS, non-patient causes, the case where patient lifestyle is a contributing factor is given 8.3% less weight (significant) in the questions based on Choice Set 2.1 and 4% less weight (insignificant) in those based on Choice Set 2.3. The case where illnesses are caused by NHS actions is significant in both comparisons, receiving 22.6% and 14.1% higher weight.

Table 5.12: Weights on condition cause/responsibility

	Choice Sets 2.1, 3.1, 4.1		Choice Sets 2.3, 3.2, 4.2	
	Mean	95% CI	Mean	95% CI
Non-NHS, non-patient causes	1.000	-	1.000	-
Partial patient causes	0.917	(0.842, 0.992)	0.960	(0.919, 1.001)
NHS causes	1.226	(1.107, 1.334)	1.141	(1.076, 1.207)

Estimates based on 5000 bootstrapped observations.
Assumes main SWF in AHYEs (Table 5.10) defines all parameters but α.
Figures based on relative marginal rates of substitution across groups.
Non-NHS, non-patient cause group is treated as baseline.

These findings are based on relatively abstract descriptions of the cause of illness and in practice these may cover a range of causes that people would have very different reactions to. In Phase 2, we compared MRSA (NHS cause), obesity (partial patient cause) and workplace exposure to hazardous substances (non-NHS, non-patient cause). In Phase 3, we replaced this last factor with a genetic condition affecting people in middle age.

Choice Sets 5.1 and 5.2 are used to assess the effect of these labels, and are in effect “labelled” versions of the more abstract Choice Sets 3.1 and 4.1. Again, the difference between the equivalent and study states for each choice set allows the identification of a change in health for both groups that yields the same social welfare. These changes can, as above, be represented in terms of societal health judgements and an average trade-off can be found. Table 5.13 presents these trade-offs for both comparisons (Choice Sets 3.1 vs 5.1 and 4.1 vs 5.1). There is generally less preference to treating obesity-related conditions over MRSA than when considering NHS caused versus partially patient caused diseases (comparing Choice Sets 3.1 and 5.1). In Phase 3, there appears to be a very similar priority the labelled case as generally given to the unlabelled case. In Phase 2, there appears to be slightly more priority given to treating workplace hazards (vs MRSA) than to treating the more generic non-patient caused diseases.

The figures here suggest that individuals would accept *some* inequality whereby those who are obese receive less lifetime health than those who suffer illness due to a genetic condition. Larger inequalities would mean that society prioritises the health of the obese group over those with genetic conditions; smaller inequalities would mean that society prioritises the group with the genetic condition. Since these effects modify the degree to which inequality is taken into account, it is inappropriate to use them in isolation of the main inequality aversion parameter. Or, in other words, the weights here are derived from the α weights alone, and not the trade-offs made in the social welfare function where health differences are considered.

Table 5.13: Equivalent health changes and trade-offs: label effects

Choice Set(s) and States	Group 1 Health	Group 2 Health	Trade-offs
Study State	66 years in 100% health 8 years in 25% health	50 years in 100% health 16 years in 25% health	
3.1 Equivalent state	(NHS causes) 59.50 years in 100% health 6.01 years in 25% health	(Non-NHS, patient) 55.01 years in 100% health 8 years in 25% health	2.46 AHYEs (NHS causes) per AHYE (Non-NHS, patient)
	-7.03 AHYEs	+ 2.86 AHYEs	
5.1 Equivalent state	(MRSA) 59.66 years in 100% health 6.65 years in 25% health	(Obesity) 55.33 years in 100% health 8 years in 25% health	2.11 AHYEs (MRSA) per AHYE (Obesity)
	- 6.70 AHYEs	+ 3.18 AHYE	
4.1 Equivalent state	(NHS causes) 59.42 years in 100% health 5.67 years in 25% health	(Non-NHS, non-patient) 54.83 years in 100% health 8 years in 25% health	2.68 AHYEs (NHS causes) per AHYE (Non-NHS, non-patient)
	-7.21 AHYEs	+ 2.69 AHYEs	
5.2 (Phase 2) Equivalent state	(MRSA) 59.30 years in 100% health 5.21 years in 25% health	(Workplace hazards) 54.61 years in 100% health 8 years in 25% health	3.20 AHYEs (MRSA) per AHYE (Workplace hazards)
	- 7.49 AHYE	+ 2.34 AHYE	
5.2 (Phase 3) Equivalent state	(MRSA) 59.35 years in 100% health 5.41 years in 25% health	(Genetic disorder) 54.70 years in 100% health 8 years in 25% health	2.72 AHYEs (MRSA) per AHYE (Genetic disorder)
	- 7.30 AHYE	+ 2.69 AHYE	

Estimates based on 5000 bootstrapped observations.
Study states defined by the choice sets.
Equivalent states derived at an aggregate level using Thurstone scores.
Trade-offs assume main SWF in order to define AHYEs differences (Table 5.10).

Note, however, that if the obesity group achieves less lifetime health then these health differences would suggest (in isolation of condition-cause/responsibility weights) they should receive *greater* priority. Beyond a critical difference in health, society would prefer to treat the obesity-related conditions because the differences in lifetime health outweighs the differences in condition cause. This was the case in Choice Set 5.1, where the overall preference gave twice the weight to treating less healthy, obese group relative to the healthier group suffering from MRSA.

5.1.2.5 Rarity

Choice Sets 6.1 and 6.2 are used to assess the effect of condition rarity. Table 5.14 outlines the equivalent health changes (the difference between the study states and equivalent states) for these two choice sets. In both choice sets, society appears to be more willing to prioritise the health of Group 2, as the less healthy group. In Choice Set 5.1, where there are no differences in rarity, a single AHYE to Group 2 is worth as much as 4.66 AHYEs to Group 1, whilst in Choice Set 6.1 where rarity is introduced, this falls to 4.00 AHYEs. This suggests that society may be more willing to prioritise groups with more rare conditions over less rare conditions. These figures relate to a question in which there is a health difference between the groups.

Where there are no health differences between the groups, then if the rare condition is given a weight of 1.00 (versus another rare condition, $\alpha = 0.5$, MRS = $\frac{\alpha}{1-\alpha} = \frac{0.5}{1-0.5} = 1$) then the extremely rare condition is given a weight of 1.19 ($\alpha = 0.543$, MRS = $\frac{\alpha}{1-\alpha} = \frac{0.543}{1-0.543} = 1.19$) versus the rare condition). The difference between these two weights is not statistically significant (95% CI = 0.785-1.597).

Table 5.14: Equivalent health changes and trade-offs: condition rarity

	Group 1 Health	Group 2 Health	Trade-offs
6.1	(Rare) +2.01 years in 100% health -20 years in 25% health	(Rare) +5.01 years in 100% health -16 years in 25% health	4.66 AHYEs (G1) per AHYE (G2)
	- 3.35 AHYE	+ 0.72 AHYE	
6.2	(Extremely rare) -6.43 years in 100% health -1.72 years in 25% health	(Rare) +5.14 years in 100% health -8 years in 25% health	4.00 AHYEs (G1, more rare) per AHYE (G2, less rare)
	- 3.26 AHYE	+ 0.81 AHYE	

Estimates based on 5000 bootstrapped observations.
Table presents differences between study state and equivalent states in each choice set.
Trade-offs assume main SWF in order to define AHYEs differences (Table 5.10).

5.1.2.6 Background characteristics and sensitivity analysis

The effect of personal characteristics was assessed by comparing the values for the parameters α, FHC and SHA (with $SHC = FHC \times SHA$). The analysis compared values for both the 12 population subgroups and between those varying over a series of other background characteristics (health, disability, insurance, health and political beliefs). The bootstrapping used 200 resamples of the data for the latter comparisons, and 5000 for the former due to some instances where estimates in the population subgroups did not lead to convergence. The tables below should be interpreted with reference to Table 5.13, presenting the main baseline results with all respondents pooled.

Background health does not appear to affect preferences either in terms of self-assessed health or self-assessed disability. Access to private health care through insurance also appeared to make no significant difference to preferences. In contrast, prior health and political beliefs do appear to affect the preferences. The four factors derived in the previous chapter (using both RI and IO data in Phase 2) were applied to Phase 2 (RI) and Phase 3 data together. The sample was then split and the data re-analysed. Table 5.15 shows the results obtained for those scoring high and low on each factor.

Those giving a higher value to “equity” are more likely to agree to statements such as “the government should redistribute income from the better-off to those who are less well-off” and “the poor are entitled to more help”. This group gives a significantly higher value to the first 25% of health, and a higher value for inequality aversion than those giving little support to equity. The “NHS” factor does not significantly affect preferences, although the findings for inequality aversion and the value of health to a child are close to significance. Agreement with this factor is associated with agreements to statements such as “Patients should receive public health care when needed and regardless of costs to the NHS” and “The government should spend more on the NHS so that it can avoid having to make difficult decisions about who gets treatment and who doesn't”, and appear to indicate a rejection of the idea of scarcity in the NHS. Overall, such individuals appear (insignificantly) less willing to prioritise (lifetime health, severity, or towards children) than those who appear more likely to accept scarcity.

Table 5.15: Effect of health and political beliefs on preferences

EQUITY		Low score (n =256)		High score (n = 255)		P
		Mean	95% CI	Mean	95% CI	
Inequality aversion	*r*	5.188	(4.334, 6.041)	7.035	(6.170, 7.901)	**0.003**
100% health as an adult	*FHA*	1.000	-	1.000	-	**-**
100% health as a child	*FHC*	1.755	(1.652, 1.858)	1.882	(1.792, 1.971)	0.069
25% health as an adult	*SHA*	0.210	(0.169, 0.251)	0.297	(0.260, 0.333)	**0.002**
25% health as a child	*SHC*	0.369	(0.283, 0.456)	0.558	(0.481, 0.635)	**0.001**

MARKET SUPPORT		Low score (n =255)		High score (n = 256)		P
		Mean	95% CI	Mean	95% CI	
Inequality aversion	*r*	6.005	(5.127, 6.883)	6.310	(5.436, 7.184)	0.630
100% health as an adult	*FHA*	1.000	-	1.000	-	-
100% health as a child	*FHC*	1.864	(1.762, 1.966)	1.794	(1.703, 1.885)	0.315
25% health as an adult	*SHA*	0.254	(0.216, 0.291)	0.263	(0.227, 0.299)	0.734
25% health as a child	*SHC*	0.474	(0.387, 0.560)	0.472	(0.396, 0.547)	0.973

MIXED		Low score (n =256)		High score (n = 255)		P
		Mean	95% CI	Mean	95% CI	
Inequality aversion	*r*	6.003	(5.198, 6.809)	6.320	(5.454, 7.187)	0.599
100% health as an adult	*FHA*	1.000	-	1.000	-	-
100% health as a child	*FHC*	1.812	(1.723, 1.901)	1.837	(1.741, 1.934)	0.708
25% health as an adult	*SHA*	0.231	(0.197, 0.266)	0.285	(0.246, 0.324)	**0.043**
25% health as a child	*SHC*	0.419	(0.346, 0.493)	0.524	(0.436, 0.612)	0.073

NHS		Low score (n =255)		High score (n = 256)		P

		Mean	95% CI	Mean	95% CI	
Inequality aversion	*r*	6.643	(5.876, 7.411)	5.583	(4.783, 6.384)	0.061
100% health as an adult	*FHA*	1.000	-	1.000	-	-
100% health as a child	*FHC*	1.888	(1.792, 1.985)	1.763	(1.670, 1.856)	0.067
25% health as an adult	*SHA*	0.271	(0.240, 0.303)	0.242	(0.205, 0.279)	0.233
25% health as a child	*SHC*	0.513	(0.440, 0.585)	0.426	(0.348, 0.505)	0.112

Results from samples of 200 bootstrapped observations.
Parameters derived using given study and equivalent states for CS2.1-2.4 for each group.

The 12 population subgroups subdivide the main sample by age, gender and education. To simplify the analysis, we use the following notation: "GxSyAz" represents these groups; for x = 1 we have males, x = 2 females, y = 1 school educated, y = 2 HE/FE educated, z = 1 under 40s, z = 2 aged 40-59, z = 3 over 60s. Our estimates allowed all inequality aversion and indifference to inequality. Convergence issues were identified with bootstrapped resamples, and in 10 of the 12 groups, all 5000 cases converged within the bootstrapping. The other groups were female, school educated groups, with under 0.1% of cases in under 40 year olds and 0.4% of cases for over 60 year olds failing to converge.

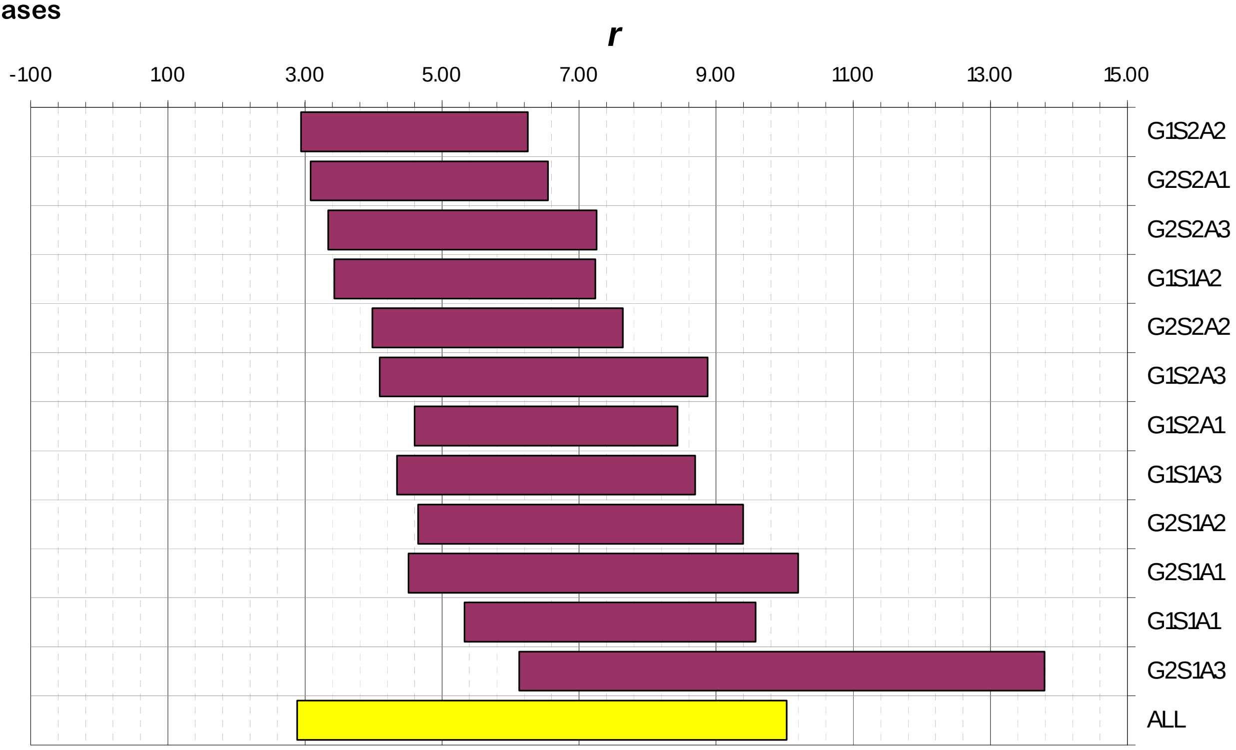
We also found a series of cases in which *r* = -1, which represents the case where individuals are indifferent to inequality; lower values are not possible without the preferences being inequality seeking. In these cases, we also find negative values below -0.250 (and above –0.40) for health at 25% in both adult and child cases. The number of these cases varies, with 8 out of the 12 groups showing at least one case, and more than 1% of cases in the school educated, female group aged 40-59 (1.2%), and the HE/FE educated groups who were female and aged under 40 (13.9%), male and aged 40-59 (2.5%), and female and aged over 60 (4.6%). We proceed by considering only these convergent, inequality neutral/inequality averse bootstrapped cases.

Figure 5.2 shows the 95% confidence intervals for the *r* parameters within these convergent, inequality averse bootstrap cases (as before, based on Choice Sets 2.1-2.4). Here, we find five significant differences in the sensitivity to inequality: the male, HE/FE educated, 40-59 group (G1S2A2) has the lowest average inequality aversion of the twelve subgroups (*r*= 4.60), and is significantly different from the two groups with the highest estimated aversion – the male, school educated, under 40 group (*r* = 7.45, G1S1A1) and the female, school educated, 60+ group (*r* = 9.96, G2S1A3). This latter group has significantly higher inequality aversion than four other groups – the two male, 40-59 groups (*r* = 4.60 and 5.33; G1S2A2 and G1S1A2); and the female, HE/FE, under 40 and 60+ groups(*r* = 6.52, *r* = 6.48; G2S2A3 and G2S2A1). As a rule, the more highly educated groups (those with "S2") appear to display less inequality aversion than the other groups.

Respondents who were female, school educated, and aged between 40 and 59 give the highest weight to childhood health (*FHC*= 2.084, G2S1A2). This value is significantly higher than both women aged under 40 group (*FHC* = 1.770 and 1.709, G2S1A1, G2S2A2), men who are HE/FE educated and aged between 40 and 59 (*FHC* = 1.608, G1S2A2), and HE/FE educated women aged over 60 (*FHC* = 1.661, G2S2A3). All groups gave values for childhood full health significantly above the adult value (*FHA* = 1.000 for all groups).

It appears that the value given to severe-ill health as an adult (*SHA*) generally increases with age. The under 40s groups, on average, give a value of 0.213 to the first 25% of health, compared to 0.251 from those aged 40-59. Those aged over 60 give a much higher average value, at 0.334. HE/FE educated women aged over 60s provide the highest value, (*SHA* = 0.431, G2S2A3), and this is significantly higher than all eight of the groups aged under 59 (male/female × schooling dimensions × two age groups). The lowest value is given by the female, HE/FE educated, under 40 group (*SHA* = 0.152, G2S2A1) and this group is significant against three of the four over 60s groups (G1S1A3, G1S2A3, G2S1A3).

Figure 5.2 Inequality aversion parameter - 95%CI for converging, inequality averse cases



As the preferences of the 12 population subgroups appears to differ, then the representativeness of the results becomes an issue. The subgroups within the existing sample of 559 individuals with full-data can be re-sampled in proportion to their population frequencies. These population frequencies based on the figures reported in Table 5.1 are given in Table 5.16, and vary from a minimum of 10 (male, HE/FE educated, 60+) to a maximum of 89 (female, school educated, 18-40 year olds). A representative sample is typically younger, and less educated than our sample (which was selected for its spread rather than its representativeness).

Table 5.16: Inferred representative sample (n = 559)

Gender	Education	Age		
		< 40	40-59	60+
Male	School only	85	66	52
	HE/FE	28	21	10
Female	School only	89	74	74
	HE/FE	29	20	11

The analysis selects the number required from each population group from the main sample and analysis proceeds as normal. Whilst our sample is unrepresentative in the pattern of respondents, the results are very similar to our earlier analyses. On none of the parameters does the difference between the original and representative samples approach significance. In general, the level of inequality aversion appears to be slightly higher (r = 6.75, 95% CI 6.19-7.32), as is the value of 100% health to a child (FHC = 1.875, 95%CI 1.817-1.933), the value of 25% health as an adult (SHA = 0.272, 95% CI 0.248-295), and 25% health as a child (SHC = 0.509, 95% CI 0.458-0.561). The condition cause/responsibility weights are also similar.

The sample used in the main result does not differentiate by data quality. Rank consistency is defined above as being satisfied where it is possible to construct a rank order of the reference and study states. There is also some weak evidence that this rank consistency also differs by population subgroup. Table 5.17 shows the proportion of rank-consistent responses by population subgroup and highlights those cases where the majority of the sample answers in a non rank-consistent fashion. As above, the consistency figures improve as we move through the each question but the baseline level of consistency differs by group. The female, school-educated 40-59 year old group has the lowest levels of rank consistency in Question 2. This may be partly behind the findings of significant differences involving this group against four others in the value of 100% health as a child. We can be less confident about these differences as a result.

Table 5.18 breaks down these rank consistency figures by choice set and interviewer, with the group with the worst rank consistency in each choice set. In over half the choice sets, the worst rank consistency was given by Interviewer H, who provided data for 20 of our 559 responses.

Table 5.17: Rank consistency by choice set and respondent subgroup

	Male						Female						ALL
	School only			HE/FE			School only			HE/FE			
	<40	40-59	60+	<40	40-59	60+	<40	40-59	60+	<40	40-59	60+	
N	44	38	43	45	50	37	45	37	56	68	56	39	559
1.1	89%	61%	74%	80%	86%	78%	76%	76%	73%	81%	84%	82%	79%
1.2	82%	74%	86%	91%	88%	92%	91%	76%	86%	84%	88%	87%	86%
1.3	84%	87%	81%	80%	88%	95%	82%	76%	84%	88%	86%	92%	85%
1.4	80%	79%	81%	82%	92%	92%	87%	84%	84%	78%	88%	79%	84%
2.1	55%	63%	51%	64%	72%	**43%**	60%	**43%**	57%	63%	55%	**49%**	57%
2.2	66%	50%	67%	69%	72%	73%	60%	**46%**	57%	72%	68%	67%	65%
2.3	77%	66%	70%	71%	82%	73%	80%	68%	66%	81%	77%	59%	73%
2.4	77%	71%	74%	62%	80%	81%	76%	68%	70%	87%	68%	69%	74%
3.1	55%	**32%**	53%	53%	72%	54%	69%	**43%**	52%	62%	57%	56%	56%
3.2	66%	53%	67%	73%	72%	78%	78%	68%	75%	81%	84%	62%	72%
4.1	**48%**	53%	70%	**47%**	76%	68%	71%	57%	**45%**	74%	73%	64%	63%
4.2	70%	58%	74%	78%	86%	76%	76%	68%	61%	88%	79%	67%	74%
5.1	55%	**42%**	58%	58%	64%	78%	**49%**	**46%**	59%	51%	61%	62%	57%
5.2	61%	55%	72%	69%	78%	68%	64%	68%	**43%**	66%	70%	51%	64%
6.1	86%	71%	84%	89%	88%	89%	89%	89%	70%	93%	88%	77%	85%
6.2	91%	89%	91%	84%	88%	92%	87%	95%	82%	87%	82%	82%	87%

AVG	71%	63%	72%	72%	80%	77%	75%	67%	66%	77%	75%	69%	72%

Individual level analysis.
Cells show the proportion of individual responses allowing rank ordering (all questions) and an inequality averse or inequality neutral position (CS1.1-1.4 only).

Table 5.18: Rank consistency by choice set and interviewer

	A	B	C	D	E	F	G	H	J	ALL
N	140	97	76	60	59	55	46	20	6	559
1.1	84%	79%	78%	87%	78%	67%	80%	**55%**	67%	79%
1.2	86%	91%	84%	87%	81%	76%	93%	**75%**	100%	86%
1.3	84%	85%	89%	83%	85%	78%	100%	**75%**	100%	85%
1.4	79%	85%	88%	88%	78%	82%	98%	**70%**	100%	84%
2.1	55%	71%	50%	63%	49%	51%	67%	**35%**	50%	57%
2.2	63%	65%	**55%**	58%	58%	71%	89%	70%	83%	65%
2.3	79%	69%	70%	72%	69%	73%	85%	**50%**	83%	73%
2.4	72%	78%	76%	70%	63%	75%	89%	**60%**	83%	74%
3.1	58%	57%	54%	50%	**39%**	47%	85%	60%	67%	56%
3.2	72%	67%	74%	**60%**	76%	73%	93%	65%	100%	72%
4.1	58%	75%	55%	62%	53%	60%	83%	60%	**50%**	63%
4.2	75%	74%	**67%**	70%	75%	67%	89%	85%	83%	74%
5.1	65%	61%	47%	53%	41%	53%	80%	**35%**	50%	57%
5.2	64%	71%	63%	**55%**	56%	60%	72%	60%	83%	64%
6.1	80%	86%	89%	88%	86%	**76%**	93%	80%	83%	85%
6.2	86%	93%	87%	92%	86%	87%	89%	**60%**	67%	87%
AVG	72%	75%	70%	71%	67%	69%	87%	62%	78%	72%

Individual level analysis.
Cells show the proportion of individual responses allowing rank ordering (all questions) and an inequality averse or inequality neutral position (CS1.1-1.4 only).

We can judge data quality for an individual by the number of choice sets where their responses are consistent with an inequality averse/inequality neutral CES SWF. By grouping this across interviewers, we can also detect whether data quality is the same across interviewers. Figure 5.3 presents the cumulative proportions of respondents who answer at least *n* (of a possible 16) choice sets in a way consistent with such a function by interviewer. Here, lower curves are preferred as they represent a more SWF consistent set of responses. Interviewer H provides the least consistent responses over a large portion of the combined graph. Note that overall (the lower dotted line) most people normally choose in a consistent fashion.

We proceed by censoring the data to remove all responses from Interviewer H as well as those respondents providing fewer than two, four, six or eight consistent choice sets. Table 5.19 presents the mean value of the main parameters across these cases. Note that the numbers excluded are relatively small at each step until we move from 6 to 8 responses, since 79 respondents had 6 or 7 consistent choice sets and are excluded at this point. Data censoring makes little difference to the parameter estimates (with the exception of the inequality aversion parameter), and with no significant differences between the cases where cut-off values are set at zero and eight (p-values all above 0.500).

Within the main analysis of the project, we assume that all health below the age of 18 is “childhood” health, and all health after the age of 18 is “adult” health. The effect of modifying this cut-off to 10 and 20 years of age is explored in Table 5.20. Overall, modifying the cut-offs makes no difference to the definition of societal health in terms of either timing or the severity of illness (*FHC*, *SHA* and *SHC*). In general, reducing the age cut-off only appears to influence the level of inequality aversion, and only very slightly. At a cut-off of 10 years, r = 5.68, whilst at 18 years, r = 6.31.

Figure 5.3: Respondent SWF consistency by interviewer

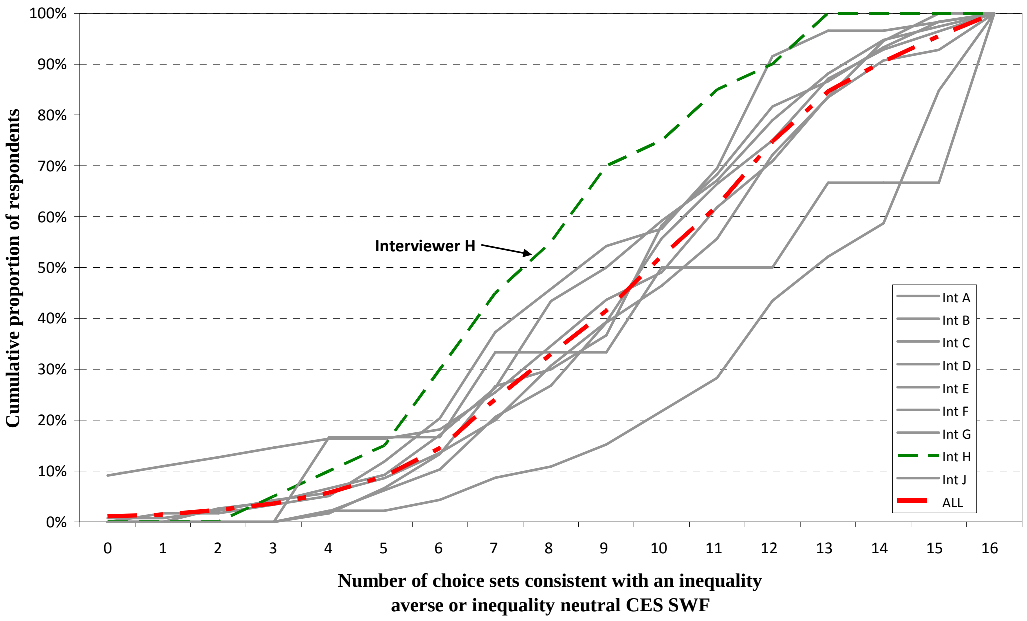


Table 5.19: Data censoring – effect of different cut-offs

		Inclusion criterion (minimum number of consistent choice sets)				
		0	2	4	6	8
	N	539	531	520	492	413
Inequality aversion	*r*	6.33	6.38	6.37	6.27	6.05
100% health as an adult	*FHA*	1.000	1.000	1.000	1.000	1.000
100% health as a child	*FHC*	1.837	1.850	1.848	1.843	1.817
25% health as an adult	*SHA*	0.273	0.268	0.269	0.266	0.263
25% health as a child	*SHC*	0.502	0.497	0.497	0.490	0.478

Results from samples of 200 bootstrapped observations.
Excludes all respondents from interviewer H
Parameters derived given study and equivalent states for CS2.1-2.4 for subgroups meeting criteria.
Cut-off provides for a minimum rationality requirement (minimum number of rank consistent choice sets) for inclusion.

Table 5.20: Effect of varying age cut-offs

		Cut off at 10 years			Cut off at 20 years		
		Mean	95% CI	P	Mean	95% CI	P
Inequality aversion	*r*	5.68	(5,16, 6.21)	0.103	6.48	(5.87, 7.10)	0.712

100% health as an adult	*FHA*	1.000	-	-	1.000	-	-
100% health as a child	*FHC*	1.825	(1.763, 1.887)	0.948	1.826	(1.763, 1.888)	0.956
25% health as an adult	*SHA*	0.268	(0.242, 0.294)	0.987	0.268	(0.242, 0.293)	0.970
25% health as a child	*SHC*	0.489	(0.434, 0.545)	0.974	0.489	(0.433, 0.544)	0.962

Results from samples of 200 bootstrapped observations. Significance versus 18 year cut-off.
Parameters derived given study and equivalent states for CS2.1-2.4 for each age cut-off.
Age cut-offs define the distinction between adult and child health. P values against standard case of an 18 year old cut-off.

The Thurstone index is used within each choice set to find an equivalent state to each study state, and hence pairs of points within the same SWF. These indices require that the data from any pairwise comparison (actual or inferred) is not unanimous, with 2% typically used as the minimum preference (98% as the maximum). Table 5.21 compares the cases where (1%, 99%) and (5%, 95%) threshold limits are instead used instead of 2% and 98%. In general, as the range of preferences allowed by the Thurstone score becomes wider (i.e. 5% to 2% to 1% cut-off values), the level of inequality aversion becomes higher (*r* = 6.24, 6.31 and 6.34) and the value of child health increases (*FHC* = 1.814, 1.828, and 1.836), as does the value of 25% health (*SHA* = 0.254, 0.268, 0.273). Since the value of 25% health to a child equals can be found by multiplying the last two items together, it also increases. However, none of the differences changes from the main case approaches statistical significance, and the precise cut-offs appear to be of relatively little importance.

Table 5.21: Effect of varying Thurstone cut-offs

		1% cut-off values		2% cut-off values (main)		P
		Mean	95% CI	Mean	95% CI	
Inequality aversion	*r*	6.34	(5.87, 6.82)	6.31	(5.44, 7.18)	0.958
100% health as an adult	*FHA*	1.000	-	1.000	-	-
100% health as a child	*FHC*	1.836	(1.785, 1.888)	1.828	(1.768, 1.888)	0.837
25% health as an adult	*SHA*	0.273	(0.253, 0.294)	0.268	(0.244, 0.292)	0.853
25% health as a child	*SHC*	0.502	(0.459, 0.546)	0.490	(0.439, 0.542)	0.640

		2% cut-off values (main)		5% cut-off values		P
		Mean	95% CI	95% CI	95% CI	
Inequality aversion	*r*	6.31	(5.44, 7.18)	6.24	(5.52, 6.96)	0.855
100% health as an adult	*FHA*	1.000	-	1.000	-	-
100% health as a child	*FHC*	1.828	(1.768, 1.888)	1.814	(1.735, 1.893)	0.782
25% health as an adult	*SHA*	0.268	(0.244, 0.292)	0.254	(0.223, 0.284)	0.634
25% health as a child	*SHC*	0.490	(0.439, 0.542)	0.460	(0.393, 0.527)	0.402

Results from samples of 200 bootstrapped observations
Parameters derived using given study and equivalent states for CS2.1-2.4.
Equivalent states defined using Thurstone scores under different cut-off values.
Standard cut-off values require aggregate proportions to fall between 2% and 98%.
Tested cut-off values require aggregate proportions to fall between 1% and 99%, or 5% and 95%.

5.2 THE ADDITIONAL STUDY

5.2.1 Methods

This study considers further the issues of timing and severity to provide further information to parameterise the SWF. It introduced illnesses at times other than at the start and end of life with illnesses affecting individuals at 10 and 30 years of age. It also considers 50% health in place of 25% health. Data was obtained in nine choice sets across three questions (A1-A3). Question A1 includes four Choice Sets (A1.1-A1.4), of which the first two are used only to give respondents some experience of the question format before the substantive data is obtained. In Question A1, all ill health occurs at the end of life. Question A2 varies the time at which ill-health occurs from the start of life (A2.1, A2.2) to 10 years of age (A2.3) and 30 years of age (A2.4). Question A3 uses a study state from the main study to allow comparison between the 25% and 50% study states.

Table 5.22 outlines the study and reference states used in the additional study. (Choice Sets A1.1 and A1.2 are not included in this table as they are used only to “warm up” respondents.) Note that Choice Sets A1.3-A1.4 and A2.1-A2.2 are identical to Choice Sets 2.1-2.4 with the 25% health states substituted for 50% health states with half the duration. Choice Sets A2.3 and A2.4 are identical to A1.3 with 8 years of illness moved from the end of life.

Table 5.22: Additional study questions

Study States	Group 1 Health	Group 2 Health
Choice Set A1.3	66 years in 100% health 4 years in 50% health	50 years in 100% health 8 years in 50% health
Choice Set A1.4	72 years in 100% health 8 years in 50% health	48 years in 100% health 8 years in 50% health
Choice Set A2.1	66 years in 100% health 8 years in 50% health	2 years in 50% health 54 years in 100% health 2 years in 50% health
Choice Set A2.2	4 years in 50% health 72 years in 100% health 4 years in 50% health	48 years in 100% health 8 years in 50% health
Choice Set A2.3	66 years in 100% health 4 years in 50% health	10 years in 100% health 8 years in 50% health 40 years in 100% health
Choice Set A2.4	66 years in 100% health 4 years in 50% health	30 years in 100% health 8 years in 50% health 20 years in 100% health
Choice Set A3.1	66 years in 100% health 8 years in 25% health	50 years in 100% health 16 years in 25% health

Reference States	Group 1 Health	Group 2 Health
x_1	62 years in 100% health 8 years in 50% health	60 years in 100% health 4 years in 50% health
x_2	60 years in 100% health 4 years in 50% health	56 years in 100% health 4 years in 50% health
x_3	59 years in 100% health 2 years in 50% health	54 years in 100% health 4 years in 50% health
x_4	58 years in 100% health	52 years in 100% health 4 years in 50% health

The analysis uses the same form of SWF as the main study. Using the inequality aversion parameter (6.32) and *FHC* parameters (1.83) from the main study, the first aim is to find the value of 50% health to an adult (*MHA*). As with the main study, the method finds the best-fit estimate that minimises the estimated errors in the utility function. This comparison uses data from A1.3-A2.2. As health and timing are assumed to be multiplicatively-related, the value of moderate health whilst a child (0-

18 years) is equal to the value of moderate health multiplied by the value of childhood versus adult health ($MHC = MHA \times FHC$).

The next questions concern the weight given to different timings. The aim here is to find the value of an adult healthy year equivalent at different points in time. These values are found (separately) for ages between 10-18 and ages 18-40 by finding the parameter values solving for equivalence between the Thurstone-equivalent and study states in Choice Sets A2.3 and A2.4. Finally, the best-fit value for SHA amongst this subgroup is found by solving for equivalence between the Thurstone-equivalent and study states in Choice Sets A3.1. The value of this parameter should be similar between Phase 2/3 and the additional study.

This analysis takes some parameters from the main study (and does not derive them separately). This may affect the variability of the estimates found and hence caution must be exercised in any comparison with the main study results. In particular, direct hypothesis tests of coefficients across the main and additional studies may be misleading. Bootstrapping is used to construct estimates of uncertainty for the additional study parameters subject to these caveats.

5.2.2 Results

5.2.2.1 Recruitment and respondents

Recruitment to the additional study took place in the two months following the main study. The additional study sampling was not designed to provide a mix of society but instead aimed to allow a quicker convenience sample which would provide indicative results only. Five of the interviewers from in the main study (Phase 3) also worked on the additional study. A total of 130 interviews were obtained in locations local to the interviewers (Cumbria, Barnsley, Sheffield and Huddersfield). Each interviewer conducted between 10 and 42 interviews. Of the 130 interviews, 129 individuals provided complete data over the choice sets A1.3-A3.1.

5.2.2.2 The relative values

The best-fit solution here suggests that 50% health is weighted as equivalent to 0.615 of an adult healthy year – this suggests that the health between death and 50% health is worth 60% more ($\frac{0.615}{0.5} / \frac{0.385}{0.5} = 1.60$) than the health between 50% health and full health. The uncertainty in this figure is found through bootstrapping (n=200), and it appears that the value given to 50% health is significantly greater than 0.50 (95% CI, 0.533-0.698). The additional study questions also investigate the value of full health years at different time periods. In the main study, ill health occurs either in the first 8 years or at the end of life, and the value of a full health year is split between the first 18 years (youth) and subsequent life (adult). The main study suggests that (lifetime equity aside) each year of full health prior to 18 years of age is worth 1.828 times as much as a year spent as an adult, with this latter figure serving as a numéraire (FHA = 1.000).

Choice Set A2.3 considers ill-health between years 10 and 18; after 18 years of age, it is assumed that the parameters for adults apply. Between 10 and 18 years, the value that minimises squared utility errors over Choice Sets A1.3-A2.3 assumes that a year of full health whilst a child is worth 1.073 times as much a year of full health as an adult (95% CI: 0.928-1.218). That is, it appears that health after the age of 10 is viewed very much like health to an adult. The analysis varying the “adult” age cut-off to 10 years above found that this made very little difference to the values of the variables defining lifetime health (the AHYE). Therefore, whilst the level for inequality aversion may change as a result of changing the cut-off, the major results of the project are unaffected.

Choice Set A2.4 considers ill-health between years 30 and 38, and was used to construct a weighting for the early period of adulthood (18 up to 40 years). For purposes of comparison, the numéraire period for health ($FHA = 1.000$) in which a year in full health equals 1 AHYE is now the period from 40 years onwards. Between 18 and 40 years, the value that minimises errors over Choice Sets A1.3-A2.2 plus 2.4 assumes that a year in full health is worth 0.989 AHYEs (95%CI: 0.888-1.091), with 50% health worth 0.609 AHYEs (95%CI: 0.499-0.720). Again, this suggests that those aged 30-38 (or by extension 18-40) are not treated differently from “normal” adults. There does not appear to be a premium placed on helping adults during the period where productivity-based estimates would suggest special emphasis is placed on health. Both estimates suggest that where lifetime health differences over an entire lifetime are accounted for (using the inequality aversion parameter), the value of health appears to be relatively constant except for a period quite close to the beginning of life, and possibly within the first 10 years.

The final choice set from the additional study used both 25% health and 50% health in the same question. The number of QALYs received in Choice Sets A1.3 and A3.1 are identical with the only difference being that Choice Set 3.1 uses 25% health for a longer period rather than 50% health for a shorter period. (The questions in Choice Set A3.1 are identical to Choice Set 2.1 in the main study.) The weight for SHA was selected here in order to solve for equality in utilities between the Thurstone-equivalent and study state in Choice Set A3.1. The best-fit figure here suggests that the first 25% of health is worth 0.231 of a full health life year (95% CI 0.212-0.249).

This estimate is lower than the estimate suggested in the main study for 25% health. Taking these figures alongside those given for 50% health, this suggests that the first 25% of health is worth 0.231 of the value of full health, with the next 25% worth 0.384 (0.615 – 0.231), and the final 50% worth 0.385 (1 – 0.615). If a standard QALY is worth £20,000, then this suggests that the health in the first 25% is worth £18,400 per QALY (£20,000 × 0.231/0.250), health in the next 25% is worth £30,720 per QALY (£20,000 × 0.384/0.250) and health in the final 50% of health is worth £15,400 (£20,000 × 0.385/0.500). This suggests that the additional study group may be placing a premium on health between 25% and 50% health.

5.3 CONCLUSION

The main study consisted of 582 interviews with members of the public, and the additional study resulted in 130 more interviews. Looking at the average preferences of the population, the main findings of Phase 3 were that people generally have preferences that diverge from simple unweighted QALY maximisation. Regarding the distribution of health across groups that differ only with respect to their lifetime health, people prefer to sacrifice overall health for a more equal distribution across those groups, rather than to have a higher level of overall health that is distributed more unequally. This is represented by an inequality aversion parameter value of $r = 6.32$. There is a premium of about 14% attached to ill health caused through the NHS compared to “non-NHS, non-patient” causes and about an 8% ‘dispremium’ attached to ill health that is partly lifestyle-related compared to “non-NHS, non-patient” causes.

In relation to the value of years of life in childhood versus adulthood, people give more value to years of life in childhood when this is defined as aged 18 or younger ($\text{FHC}_{\text{for age <18}} = 1.828$); but when this is defined as aged 10 to 18 the preference to give higher priority is no longer statistically significant from the preference given to those aged 18 and above. People give slightly more value (but not significant) to the intermediate age range 18-40 years of age, relative to aged 40 or above. Regarding severity, people regard the first 25% quality of life (of an adult) as having more value than a quarter of full health of an adult ($\text{SHA}_{\text{with 25\% QOL}} = 0.268$); and they regard the first 50% quality of life (of an adult) having more value than a half of full health of an adult ($\text{SHA}_{\text{with 50\% QOL}} = 0.615$).

A series of analyses probed into to the effect of respondent background characteristics (in terms of age, sex and education), of respondents’ health and political believes, of different interviewers, of exclusion criteria by inconsistencies, of using different Thurstone cut-offs, and of using a different definition of ‘childhood’ (i.e. <10 years as opposed to <18 years). These various analyses found that the results found in the main survey are highly robust and the impact of these considerations minimal.

CHAPTER 6: OVERALL DISCUSSION

This project builds on the existing literature on the social value of a QALY. There is a growing literature exploring whether or not publicly funded health care systems should treat all QALYs as having the same social value, and whether there is empirical support for such policies (see Dolan et al, 2005 for a review). The purpose of this project was to develop this work further and to consider whether it is possible to generate a set of equity weights for QALYs. This chapter is intended to provide a general discussion of the results from our studies.

6.1 ISSUES FOR GENERATING EQUITY WEIGHTS

6.1.1 Attributes and levels

Studies A and B in Phase 1 probed multiple attributes using the same study design, which is crucial if the results across different attributes are to be compared. The studies confirmed that age, social class, length of time with condition, dependents, quality of life without treatment, and whether the condition was caused by NHS negligence are all attributes that will make members of the public want the NHS to diverge from simple unweighted QALY maximisation. However, and encouragingly for the general framework we adopt, the studies also show that the willingness to prioritise according to an attribute is dependent on the number of QALYs that will be sacrificed in order to do. If the opportunity costs of diverging from simple QALY maximisation are not made clear, then studies may overestimate the ‘importance’ of an attribute.

Study C in Phase 1 considered the preferences of views of NHS staff. Whilst it is impractical to assume members of the public to have thought much about health care priority setting issues, NHS staff are more likely to have given the topic some thought, and by giving the NHS staff sample two explicitly differentiated questions, one on personal preference and the other on policy relevance, the study has highlighted how these may differ. Where personal views and policy relevance do not necessarily match, it is probably useful to ask them both as this may limit the degree to which one set of values are contaminated by the other (Tsuchiya et al 2003). In other words, one way to remove personal preferences from ‘policy relevance’ ones might be to ask the personal ones first to highlight that the second set of preferences should be based on something else.

Study C found that the personal and ‘policy relevance’ preferences of NHS staff are different from the social preferences of the general public (see also Dolan and Tsuchiya, 2007). The policy relevance questions in this study indicate no single attribute is relevant “no matter what”; rather, things matter “depending on what else is known”. This indicates the problematic nature of single-attribute studies, where the relevance of one attribute at a time is contrasted to simple QALY maximisation. It also warns against using the weights generated here or elsewhere at the micro or bedside level of decision-making, where more attributes will be relevant.

Of the six attributes explored in Phase 1, age, quality of life without treatment and responsibility were used in Phase 2 onwards (and in addition, rarity of condition).

Dependents, length of time with the condition and social class were dropped. The main reason for dropping these three attributes is that the complexity associated with them means that they are unlikely to be used in macro level decision-making. For example, it is not whether or not one has dependents that matters, but whether or not one is the sole responsible adult for a small child and it is not the length of time in general that is the issue, but the distinction between length of time since acquiring a condition and since being placed on a waiting list. Social class confounds several considerations such as prevalence of ill health, life expectancy, lifestyle choices, social behavioural norms, social status, financial affluence (and access to private health care) etc. Each of these considerations is linked to social class through very complex pathways. However, as we showed in the overview to the project in Chapter 1 and will consider further below, it is possible from our results to infer social class weights attributable to inequalities in life expectancy.

6.1.2 Concentration and dispersion

The final study in Phase 1, Study D, was based on a quasi-replication of earlier study by Rodriguez and Pinto (2000). If the health benefit per individual patient is large enough, then people will prefer to disperse health than to concentrate them. However, dispersion is not supported if the benefit per head becomes too small. The ‘tipping point’ of about 2.6 years is of direct relevance to studies focusing on health gains rather than outcomes as in our SWF approach. However, it was relevant in designing the main preference elicitation task to determine the inequality aversion parameter, since if the inequalities in outcomes are too small (i.e. less than 2.6 years), respondents may not be willing to sacrifice efficiency in order to ameliorate it.

6.1.3 The social welfare function

Following other research within health economics, we use a social welfare function (SWF) that allows trades-off between the lifetime health of different individuals. Since society may give different priorities to the health received at different ages and to different levels of severity, we consider social valuations over lifetime health and in relation to what we refer to as ‘adult healthy year equivalents’ (AHYEs) rather than in relation to ‘standard’ QALYs. In keeping with the conventional SWF framework, we focus on the valuation of outcomes rather than gains. We believe this focus on outcomes is preferable given its generalisability, as well as empirical concerns over the formation of reference points and violations of the Pareto principle (that is, from a given reference point, preferring to do less for both groups).

To our knowledge, Nord et al (1999) was only previous paper to have advocated social judgements over social valuations of health. There, social valuations were limited to the somewhat unrealistic case that society will value every year of life lived to all chronically ill or disabled individuals equally so long as the individual concerned would prefer this state to death. This assumption was made to address the conflation of social values with the measurement of individual health benefits in person trade-off data. We believe our methods and analyses are unique in estimating both the social valuation of lifetime health and the value given to more equal lifetime health together.

6.1.4 Design and analysis

To generate stable preferences, we recommend the use of attitudinal questions regarding health and political beliefs prior to the main quantitative elicitation tasks. However, due to the fact that their impact is uncertain and likely to be rather temporary, we do not recommend the use of resource intensive group discussions to precede the individual quantitative exercise. There are also logistic and practical constraints with running discussion groups and the opportunity costs are very high in terms of sample sizes and resources.

This project has used individual level ordinal data and analysed it as cardinal information at the aggregate level, using the Thurstone approach. Given that the profiles and distributions of health that respondents were asked to consider were not simple, it made sense to make the elicitation task as simple as possible, so a pairwise choice format was used. One challenge was how to reflect uncertainty in the results. The uncertainties included in this report are obtained by bootstrapping, which were computationally very time consuming.

It is often pointed out that issues around health care priority-setting are not something members of the public routinely think about, and thus in order to give answers that they can feel confident about, respondents should be allowed more opportunity to digest and to reflect on the relevant issues before committing themselves to a preference. In order to address this concern, the main preference elicitation task was preceded by a set of health and policy belief questions, unpacking people’s attitudes towards priority setting in health care. Since this was a relatively low cost activity to include, which may make it easier for respondents to think about their preferences, the attitudinal questions were a non-negotiable part of the study.

More contentious was the possible impact of including group discussion sessions, to allow participants to exchange and challenge their views, and a deliberation period, between the discussion and the individual interview. Phase 2 compared the two designs. Although the prior anticipation was that the two designs will result in different preferences, the results remained inconclusive. However, one clear finding from the Phase 2 study was that the resource intensive design is not practical, given the logistic complication of coordinating groups of participants.

6.2 THE MAIN RESULTS

Taken together, the main study and the additional study interviewed over 700 members of the public. The main findings of the study are s Table 6.1 below.

6.2.1 Timing and severity

The timing of ill health seems to have a substantial impact. A greater weight was given to health experienced as a child in the main study (between ages 0-18 versus 18+), and the additional study suggesting that this might only apply within the first 10 years of life. We re-ran our main analysis to compare health in the first 10 years versus other timings (i.e. a 10 year cut-off for “childhood”) and found no significant differences in our SWF parameters. This supports the use of a higher cost-per-QALY threshold for treatments that involve young patients. However, those interpreting the results of this study should refrain from extrapolating values in ranges that lie beyond the values used in the elicitation task. For example, in Choice Sets 2.1-2.4, where the impact of the timing of ill health is explored, none of the states involve a person dying before the age of 18.

Table 6.1: Summary of results from Phases 2 and 3
Baseline threshold of £20,000 for one year in full health as an adult (i.e. one AHYE)

Issue	Study	Results	Threshold per QALY change**
Timing of health	M	Significantly higher weight is given to health under 18 as compared to over 18.	£36,560 for age < 18, £20,000 for age > 18*
	A	There is no significant difference in the weight given to 10-18 years versus 18+ years, and no difference in the weight given to 18-40 years versus 40+ years.	£36,560 age < 10*, £21,460 age 10-18, £19,780 age 18-40 £20,000 age > 40*
Severity	M	The first 25% of health is given more weight than predicted by the QALY but not significantly so	£21,440 (4 years in 25% health) £19.520 (1.33 years from 25% to 100%)
	A	The first 50% of health is given significantly more weight than predicted by the QALY, with a premium between 25% and 50%.	£18,480 (4 years in 25% health) £30,720 (4 years from 25% to 50%) £15,400 (2 years from 50% to 100%)
	I	Premium on 25%-50% health remains but is reduced when considering 50% estimate from additional study with 25% estimate in main study	£21,440 (4 years in 25% health) £26,320 (4 years from 25% to 50%) £15,400 (2 years from 50% to 100%)

Issue	Results	Threshold for health change (per AHYE)
Inequality aversion	Significant difference from the inequality neutral case (*r* = -1), where health is summed across individuals i.e. willingness to prioritise more equal outcomes. Indicative weights can be calculated by social class based on life expectancy at birth 2002-2005. Figures are approximate, ignore any potential social mobility and any weighting given to responsibility	£17,530 per AHYE (Class I) £18,450 per AHYE (Class II) £19,130 per AHYE (Class IIIN) £20,670 per AHYE (Class IIIM) £21,160 per AHYE (Class IV) £23,120 per AHYE (Class V)
Condition cause	Significantly higher weight to NHS causes and significantly lower weight to partial patient lifestyle causes versus non-patient, non-NHS causes	£18,340 per AHYE (partial patient cause) £24,520 per AHYE (NHS cause) £20,000 per AHYE (other causes)
Rarity	Extremely rare conditions receive a higher weight than rare conditions but this is not significant	£20,000 per AHYE (rare condition)* £23,760 per AHYE (extremely rare)

Main (M), Additional (A), Indirect from main and additional studies (I)

* Fixed by assumption.

** Timing changes assume moving an individual from 0% to 100% health, severity changes assume an improvement to adult health.

In relation to severity, 25% health was valued as worth 0.268 of full health, so that there was a slight premium in the first 25% of health (but not significantly more than the 0.25 weight given by the standard QALY). 50% health was valued as worth 0.615 of full health, so that again there was a premium given to lower health states versus higher health states. Comparing 25% health and 50% health in the additional study, 25% health was valued as worth 0.231 of full health, with the health between 25% and 50% health therefore worth 0.384 (0.615-0.231). This suggests a slight premium for the health between 25% and 50% health. An indirect comparison across the main and additional studies suggests that the first 25% is worth 0.268, the next 25% is worth 0.347 (0.615-0.268) and the final 50% is worth 0.385.

6.2.2 Inequality aversion and responsibility: the example of social class

Overall, respondents were willing to make trade-offs between total health and a more equal distribution of health. The inequality aversion parameter can be used to estimate the social value of an AHYE across a range of characteristics foe which data can be generated. As an illustrative example, consider the case of social class. Official statistics[3] suggest that the life expectancy for those born between 2002 and 2005 will continue to vary by social class. Combining male and female data using the rate of male and female births in 2005[4], those in social class I will live, on average, for 82.5 years and those in social class V will live, on average, for 75.3 years. As an illustration and for simplicity, we ignore the impact of illness and assume only years in full health. We examine the social welfare of an additional AHYE to each of the six social classes (i.e. treating IIIN (non-manual) and IIIM (manual) separately).

For each social class, we calculate the social welfare value of an additional AHYE when compared to each other social class. Table 6.2 shows the ratios that are generated: For example, an extra AHYE to those in the more disadvantaged social class V is worth 32% more than a AHYE to those in the social class I. If cost-effectiveness thresholds have a mean value of £20,000 per adult healthy years-equivalent (AHYE) across these groups, then these figures are consistent with the following class-specific thresholds: £17,530 for I; £18,450 for II; £19,130 for IIIN; £20,670 for IIIM, £21,160 for IV); and £23,120 for V. These figures are indicative, and could change for a variety of reasons including where: illness is incorporated, social mobility is incorporated, and the causes of the different life expectancies (lifestyle etc.) are fully accounted for.

Table 6.2: implied weights for inequality aversion by social class

	Life expectancy	Trade-offs					
		V	IV	IIIM	IIIN	II	I
V	75.3	1.0000	1.0924	1.1219	1.2086	1.2534	1.3187
IV	77.7	0.9154	1.0000	1.0271	1.1064	1.1474	1.2072
IIIM	78.4	0.8913	0.9736	1.0000	1.0772	1.1172	1.1754
IIIN	80.3	0.8274	0.9038	0.9283	1.0000	1.0371	1.0911
II	81.3	0.7978	0.8715	0.8951	0.9643	1.0000	1.0521
I	82.5	0.7583	0.8284	0.8508	0.9165	0.9505	1.0000

Our findings suggest that respondents were willing to give different weight to groups according to the cause of ill health. Our estimates suggest a 14% higher weight to treating NHS-caused illnesses when compared to “non-NHS, non-patient” caused illnesses (in the case where lifetime health is identical). Those whose illnesses are

---

[3] Trends in Life Expectancy by social class 1972-2005: http:/www.statistics.gov.uk/downloads/theme_population/Life_Expect_Social_class_1972-05/life_expect_social_class.pdf. [Accessed Jan 2008]

[4] Birth Statistics: http://www.statistics.gov.uk/downloads/theme_population/FM1_34/fm1_no34_2005.pdf

partially caused by their lifestyle get 8% less weight than those in which illness is "non-NHS, non-patient" responsibility.

Whether or not any of these groups would actually receive more or less weight will depend on the direction and the size of the differences in health that exist between. Consider two groups who have identical health and a lifestyle which does not contribute to their chance of developing illness. Now suppose one of these groups changes its behaviour and is more likely to develop an illness – this group's "cause" has changed and its future health has worsened. Whilst the condition cause weight suggests that the group contributing to their illness receives less priority, the new health inequalities will suggest that they will receive greater priority.

In fact, the differences do not have to be that large before the health inequality dominates the condition cause weight. Consider groups who would live for 70 years in full health before dying. It would only take one group to lose just over one year of life as a result of their lifestyle for them to be given greater priority. To place this in context, we might expect lifestyle-related illness to impact differently on those in lower social classes. Consider the extreme case where those in social class I do not suffer from such "self-inflicted" illness, and that such illnesses account for all the health differences between social class I and the other social classes. For social classes II through V, the health inequalities are all of sufficient magnitude that they would dominate responsibility and hence receive greater priority than social class I.

Therefore, regardless of why these health inequalities occur, the results from our study suggest that that society will generally prefer to prioritise social classes II to V over social class I. However, one potentially very important finding for health policy is that 'labels' make a big difference. For example, when a condition that is a function of genes, environment and lifestyle is instead labelled as 'obesity', the respondents appear to latch on to the lifestyle bit and give about 25% less weight to an AHYE to an obese group as compared to a non-obese group.

More generally, respondents may latch onto salient characteristics of other groups, such as gender and race, so it may well be more appropriate to elicit weights for inequality and responsibility in more abstract ways and then apply them to particular groups, in much the same way as values for somewhat abstract health states (e.g. defined in terms of the EQ-5D) are elicited and then applied to condition 'labels', such as cancer, stroke and heart disease.

6.2.3 Rarity

Rarity of the condition was introduced at Phase 2, and as a result, we do not have detailed qualitative evidence to suggest why this should be regarded as a reason for diverging from simple QALY maximisation, nor do we have any evidence regarding the views of NHS staff. The results from the main study suggest that there is a preference towards giving higher priority to extremely rare conditions compared to rare conditions of around 20% but this does not reach significance due to a much higher level of uncertainty in this estimate than elsewhere.

6.2.4 The effect of background characteristics and sensitivity analyses

Within this project we did not aim for representative preferences within our main estimates, but instead investigated whether a representative mix of our sample would have made different choices. Whilst we found some evidence that preferences differed by gender, age, and education, we did not find any significant differences with a representative mix as compared to our main sample. Modifying our sample by including only those with a minimum number of “consistent” responses (and excluding responses from one interviewer) had no significant effect on the results, and neither did varying the cut-offs used for proportions when computing Thurstone scores.

6.3 SOME METHODOLOGICAL LESSONS AND LIMITATIONS

Overall, our aims have been met: we have identified relevant attributes and levels, we have used attitudinal questions to elicit more stable preferences and we have estimated equity weights based on the preferences of the general public. We have learned a number of lessons through this research and there are, of course, some limitations to what we have done.

6.3.1 Lessons

This project has made several methodological contributions to the measurement of social welfare functions. Like discrete choice experiments (DCEs), the methodology is based on a random utility model. Whilst DCEs are well-established, they are limited in the degree to which they can include non-linearities in their functional form and, as such, they are best placed to capturing preferences in a SWF. The method developed here uses a random utility measure in the form of Thurstone scores to identify points that are equivalent in social welfare terms. These points are then used to parameterise a SWF.

Significantly, the parameterisation of the SWF allows both the definition of what counts as "lifetime health" and the sensitivity to differences in lifetime health to be assessed together. Previous methods typically consider only one or other of these options. Notably, person trade-off type research assumes a sum-ranking SWF and investigates the value of (lifetime) health, whilst William's fair innings argument assumes the QALY as a measure of lifetime health, and investigates how sensitive individuals are to differences in lifetime QALYs. Neither method is sufficient when individuals simultaneously weight lifetime health differently to the QALY model and care about inequalities

Overall, the results found here could be interpreted as a proof of concept more than a definitive answer as to what matters to the public, and how much those thing matter. The SWF approach, which has been around a long time as a conceptual tool in economics, would also appear to be empirically feasible and it therefore represents a potentially powerful way to analyse distributional issues in health and elsewhere.

6.3.2 Limitations

The project has considered a very complex topic which is likely to have been challenging for many participants. Our study may have benefited from some qualitative study within Phases 2 and 3, and specifically a formal “think aloud” approach during piloting as compared to the informal feedback we received from interviewers. We would also have gained much in Phase 2 from recording the discussion groups but transcribing and analysing such data, as past experience has shown, is a particularly time-consuming activity.

More generally, the questionnaire in Phases 2 and 3 required a balance between breadth of approach and triangulation. The issues raised by the study are complex and the main questionnaire required a lengthy book of prompts (74 pages) and interview (53 pages). In an ideal world, our main SWF would be based on a larger number of choice sets and would consider different amounts of lifetime health. We would ideally have liked more levels for our attributes, particularly in the severity dimension. However, this was not practical given the size of the questionnaire. Overall, we believe that our design balances respondent fatigue and comprehensiveness of our results as well as we could be expected of the project.

So far as has been possible within time and resource constraints, we have addressed potential weaknesses in the methods within the sensitivity analysis and additional studies. Ideally, we would have been able to consider results from a wider variety of methods in both the elicitation and in the subsequent analysis of data. This was, however, beyond what could be achieved within the project timeframe but could of course be explored in future work. The results sections discuss the outcomes in terms of statistically significant differences in the key parameters but, at a practical level, what ultimately matters is whether the results would make a difference in terms of the incremental cost-effectiveness of an intervention. This is a complex issue that is beyond the aims of this project, and thus the report does not address this.

The results that generate the SWF should always be placed in the appropriate context; that is, in the context of macro level decisions and not micro level ones. To illustrate this, the inequality aversion parameter suggests that an additional year of life in full health for a 20 year-old who is about to die is worth about 1500 times as much as the same benefit to an 80 year-old. However, if the 20 year old and 80 year old were instead in groups of 100 identical individuals who would each live for 70 years in full health before dying, then the relative weights to the group containing the 20 year-old falls to 1.01. Naturally, as the size of the reference group increases, the importance of individual cases falls and it is the general relationship in the health received across these very large groups which are of interest.

At a macro level, however, we look forward to all of our results contributing to on-going debate and health policy in relation to the relative societal value of health gains to different beneficiaries.

REFERENCES

Abásolo I., Tsuchiya A (2004). Exploring Social Welfare Functions and violation of monotonicity: An example from inequalities in health. *Journal of Health Economics* 23(2): 313-329.

Atkinson AB (1970). On the measurement of inequality, *Journal of Economic Theory* 2:244-263.

Banister P, Burman E, Parker I., Taylor M, Tindall C (1994). *Qualitative methods in psychology: a research guide*. Buckingham, England: Open University Press.

British Household Panel Survey (2006). Volume B - The codebook. http://www.iser.essex.ac.uk/ulsc/bhps/doc/volb/ [Accessed Jan 2008].

Bleichrodt H, Diecidue E, Quiggin J (2004). Equity weights in the allocation of health care: the rank-dependent QALY model. *Journal of Health Economics* 23:157-71.

Bommier A, Stecklov G (2002). Defining health inequality: why Rawls succeeds where social welfare theory fails. *Journal of Health Economics* 21(3):497-513.

Bray RM, and Noble AM (1978). Authoritarianism and decisions of mock juries: Evdience of jury bias and group polarization. *Journal of Personality and Social Psychology* 36:1424-30.

Browning CJ, Thomas SA (2001). Community values and preferences in transplantation organ allocation decisions. *Social Science & Medicine* 52(6):853-61.

Choudhry N, Slaughter P, Sykora K, Naylor D (1997). Distributional dilemmas in health policy: large benefits for a few or smaller benefits for many? *Journal of Health Services Research and Policy* 2:212-216.

Cowell FA, Amiel Y, Polovin A (2001) Risk perceptions, income transformations and inequality. *European Economic Review* 45:941-952.

DeSantis L, Ugarriza D (2000). The concept of theme as used in qualitative nursing research. *Western Journal of Nursing Research* 22(3):351-372.

Dolan P, Robinson A (2001). The measurement of preferences over the distribution of benefits: The importance of the reference point. *European Economic Review* 45(9):1697-1709.

Dolan P, Shaw R, Tsuchiya A, Williams A (2005). QALY maximisation and people’s preferences: A methodological review of the literature. *Health Economics* 14(2):197-208.

Dolan P (1998). The measurement of individual utility and social welfare. *Journal of Health Economics* 17(1):39-52.

Dolan P (2000). “The measurement of health-related quality of life for use in resource allocation decisions in health care” in AJ Culyer and JP Newhouse (eds), *Handbook of Health Economics* pp. 1723-1760.

Dolan P, Green C (1998). Using the person trade-off approach to examine differences between individual and social values. *Health Economics* 7(4):307-12.

Dolan P, Olsen JA (2001). Equity in health: the importance of different health streams. *Journal of Health Economics* 20:823-834.

Dolan P, Tsuchiya A (2003). *The social welfare function and individual responsibility: Some theoretical issues and empirical evidence from health*. Sheffield Health Economics Group, University of Sheffield, Discussion Paper Series 03/1.

Dolan P, Tsuchiya A (2007). Do NHS staff and members of the public share the same views about how to distribute health benefits? *Social Science & Medicine*, 64: 2499–2503

Dolan P, Cookson R, Ferguson B (1999). Effect of discussion and deliberation on the public's views of priority setting in health care: focus group study. *British Medical Journal* 318(7188):916-9.

Dolan P, Tsuchiya A, Smith P, Shaw R, Williams A (2002). *Determining the parameters in a social welfare function using stated preference data: An application to health.* Sheffield Health Economics Group, University of Sheffield, Discussion Paper Series 02/2.

Dolan P, Olsen JA, Menzel P and Richardson J, An inquiry into the different perspectives that can be used to elicit preferences to inform priority-setting, *Health Economics*, 12, 7, 545-551, 2003.

Dutta I (2007). Health inequality and non-monotonicity of the health related social welfare function. *Journal of Health Economics* 26(2) 414-31.

Edwards AL (1957) *Techniques of attitude scale construction*. New York: Appleton-Century- Crofts.

Edwards RT, Boland A, Wilkinson C, Cohen D, Williams J. (1999). Choosing explicit criteria for the *prioritisation of elective NHS waiting lists : survey evidence of clinical and lay preferences from Wales*. Paper presented at the Health Economists Study Group Meeting

Emmelin MAC, Lindholm LA, Stenlund HC, Dahlgren LG (1999). Pol-ethical considerations in public health - The views of Swedish health care politicians. *European Journal of Public Health* 9(2):124-30.

Guilford, JP (1954). *Psychometric methods*. New York: McGraw-Hill.

Gyrd-Hansen D (2004). Investigating the social value of health changes. *Journal of Health Economics* 23:1101–1116.

NICE (2004). Guide to the Methods of Technology Appraisal (reference N0515). http://www.nice.org.uk/niceMedia/pdf/TAP_Methods.pdf [Accessed Jan 2008].

Isenberg D (1986). Group Polarization: A Critical Review and Meta-analysis. *Journal of Personality and Social Psychology* 50(6):1141-51.

Kahneman D, Tversky A (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263-291.

Kvale S (1996). An introduction to qualitative research interviewing. Thousand Oaks, CA, US: Sage Publications, Inc.

Layard R, Walters AA (1994). Allowing for income distribution. In *Cost-Benefit Analysis* (Layard R, Glaister S. eds) Cambridge University Press (reprinted from Layard R, Walters AA, 1978, *Microeconomic Theory*, McGraw-Hill).

Lerner MJ (1980). *The belief in a just world: A fundamental delusion*. New York: Plenum Press.

Lindholm L, Rosen M. (1998). On the measurement of the nation's equity adjusted health. *Health Economics* 7(7):621-8.

Little IMD, Mirrlees JA (1974). *Project Appraisal and Planning for Developing Countries,* Heinemann Educational Books.

Lloyd A (2003). Threats to the estimation of benefit: Are preference elicitation methods accurate? *Health Economics* 12(5):393-402

Louviere J, Hensher D, Swait J (2000). *Stated Choice Methods: analysis and application.* Cambridge: Cambridge University Press

McCabe C, Claxton K, Tsuchiya A (2005). Orphan Drugs and the National Health Service: Should we value rarity? *British Medical Journal*, 331(7523):1016-1019

Moscovici S, Zavalloni M (1969). The group as a polarizer of attitudes. *Journal of Personality and Social Psychology* 12(2):125-35.

Menzel P (1999). How should what economists call ‘social values’ be measured? *The Journal of Ethics* 3:249-273.

Nord E (1995). The Person-Trade-Off Approach to Valuing Health-Care Programs. *Medical Decision Making* 15(3):201-8.

Neuberger J, Adams D, MacMaster P, Maidment A, Speed M. (1998). Assessing priorities for allocation of donor liver grafts: survey of public and clinicians. *British Medical Journal* 317(7152):172-5.

Nord E, Street A, Richardson J, Kuhse H, Singer P (1996). The significance of age and duration of effect in social evaluation of health care. *Health Care Analysis* 4(2):103-11

Olsen JA (2000). A note on eliciting distributive preferences for health. *Journal of Health Economics* 19(4):541-50.

Payne J, Bettman J, Johnson E (1988). Adaptive Strategy Selection in Decision Making. *Journal of Experimental Psychology* 14:534-52.

Ratcliffe J (2000). Public preferences for the allocation of donor liver grafts for transplantation. *Health Economics* 9(2):137-148

Ratcliffe J, Bekker HL, Dolan P, Edlin R (2007). *The trade-off between access, equity and cost-effectiveness: integrating preferences and attitudes.* Unpublished Paper.

Rodriguez E, Pinto JL. (2000). The social value of health programmes: Is age a relevant factor? *Health Economics* 9(7):611-21

Rodriguez-Miguez E, Pinto-Prades JL (2002). Measuring the social importance of concentration or dispersion of individual health benefits. *Health Economics* 11:43-53.

Roberts T, Bryan S, Heinbotham C (1997). Public involvement in health care priority setting: an economic perspective. *Health Expectations* 2:235-44.

Seale C (2000). *The quality of qualitative research*: *Introducing qualitative methods.* Thousand Oaks, CA, US: Sage Publications, Inc, 214.

Stalmeier PFM, Bezembinder TGG, Unic IJ (1996). Proportional heuristics in time tradeoff and conjoint measurement. *Medical Decision Making* 16(1):36-44.

Stewart DW, Shamdasani PN (1990). *Duty Bound: Elder Abuse and Family Care.* Newbury Park, CA: Sage.

Stoner JAF (1961). *A comparison of individual and group decisions involving risk. Unpublished master’s thesis.* School of Industrial Management, Massachusetts Institute of Technology.

Sutherland HJ, Llewellyn-Thomas H, Boyd NF, Till JE. (1982) Attitude toward quality of survival: the concept of maximal endurable time. *Medical Decision Making* 2:299-309.

Tajfel H (ed) (1982). *Social identity and intergroup relations*. Cambridge: Cambridge University Press.

Thurstone LL (1927a). A law of comparative judgment. *Psychological Review*, 34, 273-286.

Thurstone LL (1927b). The method of paired comparisons for social values. *Journal of Abnormal and Social Psychology*, 21, 384-400.

Tsuchiya A (2001). *The value of health at different ages*. Centre for Health Economics, University of York. Discussion Paper 184

Tsuchiya A, Dolan P, Shaw R (2003). Measuring people's preferences regarding ageism in health: Some methodological issues and some fresh evidence, *Social Science and Medicine* 57:687-696.

Ubel PA (1999). How stable are people's preferences for giving priority to severely ill patients? *Social Science & Medicine* 49(7):895-903.

Ubel PA, Baron J, Nash B, Asch DA (2000). Are preferences for equity over efficiency in health care allocation "all or nothing"? *Medical Care* 38(4):366-73.

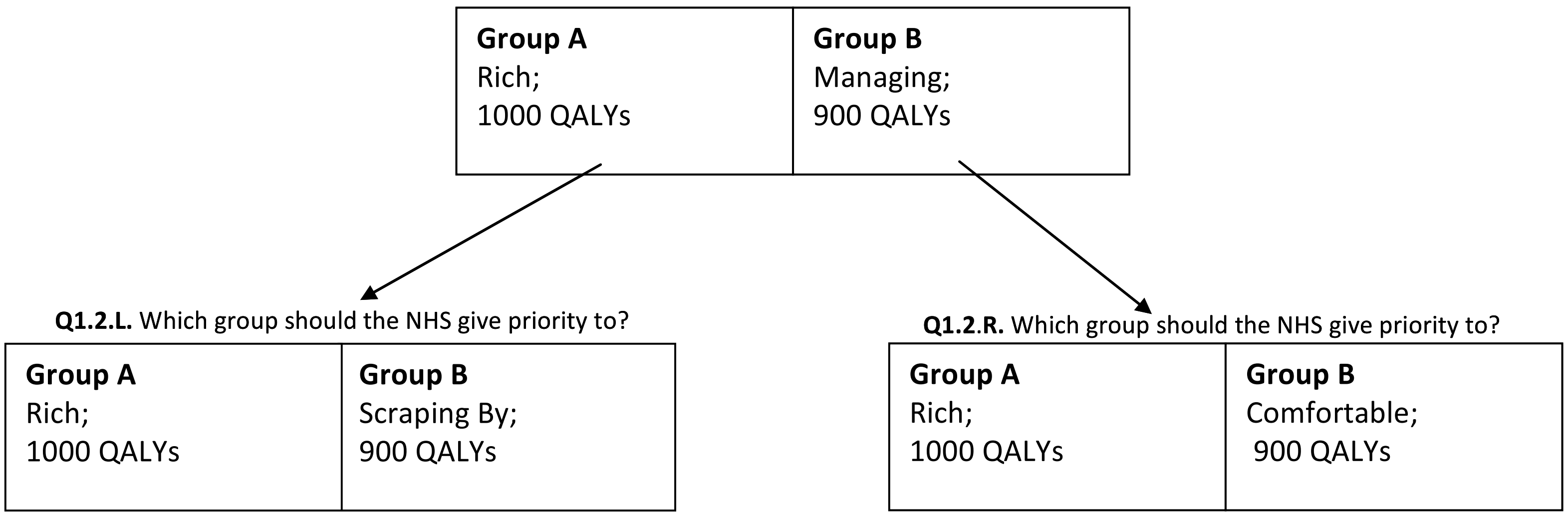
Ubel PA, Dekay ML, Baron J, Asch DA. (1996) Cost-effectiveness analysis in a setting of budget constraints - Is it equitable? *New England Journal of Medicine* 334(18):1174-7.

Ubel PA, Jepson C, Baron J, Mohr T, McMorrow S, Asch DA. (2001). Allocation of transplantable organs: Do people want to punish patients for causing their illness? *Liver Transplantation* 7(7):600-7.

APPENDIX 1: PHASE 1, STUDY B SAMPLE QUESTION

Key:

Rich	**-** Upper 25% of population
Comfortable	**-** Next 25% of population
Managing	- Next 25% of population
Scraping by	**-** Next 25% of population



Group A Rich; 1000 QALYs	**Group B** Managing; 900 QALYs

Q1.2.L. Which group should the NHS give priority to?

Group A Rich; 1000 QALYs	**Group B** Scraping By; 900 QALYs

Q1.2.R. Which group should the NHS give priority to?

Group A Rich; 1000 QALYs	**Group B** Comfortable; 900 QALYs

APPENDIX 2: PHASE 1, STUDY C POSTAL QUESTIONNAIRE

Section C. Cost effectiveness and patient characteristics

One way to think about cost effectiveness is by giving equal weight to a unit of health (e.g. QALYs) irrespective of patient characteristics. But actual policy making may need to take into account additional considerations.

Question1.
Questions 1 is concerned with your personal preferences, i.e. your private views. Imagine that there are two programmes (programme A and programme B) of equal cost effectiveness and you have to select one programme to prioritise. How much would you be personally be inclined to take each of the following patient characteristics into account in decision making at policy level (not at the bedside level).

In making a choice about which program to prioritise, how much would you be personally inclined to take each of the following patients characteristics into account: Note: All programmes are of equal cost effectiveness	I'd definitely prioritise programme A, no matter what	I'd probably prioritise programme A depending on what else is known about the patients	I'd have no preference between the programmes	I'd probably prioritise programme B depending on what else is known about the patients	I'd definitely prioritise programme B, no matter what
1. Age: Programme A applies to patients aged 5-25 and programme B applies to patients aged 60-80					
2. Dependents: Programme A applies to patients without dependents and programme B applies to patients with dependents					
3. Length of time with condition: Programme A applies to patients who have had the condition for 1 year and programme B applies to patients who have recently got the condition					
4. Quality of life without treatment: Programme A applies to patients whose quality of life is 70% but with treatment will be increased to 90% and programme B applies to patients whose quality of life is 40% but with treatment will be increased to 60%					
5. NHS negligence: Programme A applies to patients whose condition was caused by NHS negligence and programme B applies to whose condition was caused by natural causes					
6. Social class: Programme A applies to patients who are in social class 1 (e.g. doctors, lawyers and clergy) programme B applies to patients who are in social class 5 (e.g. couriers, labourers and refuse collectors					

Question 2.

Question 2 is concerned with how relevant you think the same patient characteristics would be in the decision-making at policy level (not at the bedside level), i.e. question 2 is not on what your personal preferences are, but rather about what policy could take into account. Please detach yourself from what you personally would prefer, to what you think is right in the NHS to apply as guiding principles.

Imagine that there are two programmes of equal cost effectiveness and policy makers have to select one programme to prioritise. How relevant do you think each of the following patient characteristics would be in decision making at policy level (not the bedside level). These questions are not on what your personal preferences are but rather about what policy would take account of.

In making a choice about which program to prioritise, how relevant do you think the following patient characteristics would be in the decision to prioritise either programme at policy level: Note: These questions are not on you personal preferences but about what policy could take account of; and all programmes are of equal cost effectiveness	This characteristic would be relevant no matter what	This characteristic would be relevant depending on what else is known about the patients	This characteristic would not be relevant at all	Policy maker would not want to know this information
1. Age: Programme A applies to patients aged 60 to 80 and programme B applies to patients aged 5 to 25				
2. Dependents: Programme A applies to patients with dependents and programme B applies to patients without dependents				
3. Length of time with condition: Programme A applies to patients who have recently got the condition and programme B applies to patients who have had the condition for 1 year				
4. Quality of life without treatment: Programme A applies to patients whose quality of life is 40% but with treatment will be increased to 60% and programme B applies to patients whose quality of life is 70% but with treatment will be increased to 90%				
5. NHS negligence: Programme A applies to patients whose condition was caused by NHS negligence and programme B applies to patients whose condition was caused by natural causes				
6. Social class: Programme A applies to patients who are in social class 1 (e.g. doctors, lawyers and clergy) programme B applies to patients who are in social class 5 (e.g. couriers, labourers and refuse collectors				

7. If you think that some of the above attributes would be relevant at policy level but feel that they would not actually be **applied** could you please state below the reason(s) why?

APPENDIX 3: PHASE 1, STUDY D POSTAL QUESTIONNAIRE

Example of a set of cards used in Study D (not to scale)

Card K		Card S	
Treatment A	Treatment B	Treatment A	Treatment B
40 years	20 years	40 years	20 years
½ QOL	½ QOL	½ QOL	½ QOL
5 people	1 person	5 people	20 people
Card G		Card P	
Treatment A	Treatment B	Treatment A	Treatment B
40 years	20 years	40 years	20 years
½ QOL	½ QOL	½ QOL	½ QOL
5 people	3 people	5 people	18 people
Card Q		Card T	
Treatment A	Treatment B	Treatment A	Treatment B
40 years	20 years	40 years	20 years
½ QOL	½ QOL	½ QOL	½ QOL
5 people	5 people	5 people	15 people
Card M		Card U	
Treatment A	Treatment B	Treatment A	Treatment B
40 years	20 years	40 years	20 years
½ QOL	½ QOL	½ QOL	½ QOL
5 people	8 people	5 people	12 people

APPENDIX 4: PHASE 2/3 PROMPTS

The interview only version of the questionnaire was used in both Phase 2 and Phase 3 is attached to this report. There are some very slight differences between the two phases (one label differs in four choices), and it is the Phase 3 version that is given here.