EEPRU

Policy Research Unit in
Economic Evaluation of Health
and Social Care Interventions

Research Report

Title: Eliciting societal preferences for burden of illness, therapeutic improvement and end of life for value based pricing: a report of the main survey

Authors: John Brazier, Donna Rowen, Clara Mukuria, Sophie Whyte, Anju Keetharuth, Arne Rise Hole, Aki Tsuchiya, Phil Shackley

No: 011

Date: October 2013

The University Of Sheffield.

THE UNIVERSITY of York

The Policy Research Unit in Economic Evaluation of Health and Care interventions is funded by the Department of Health Policy Research Programme. It is a collaboration between researchers from the University of Sheffield and the University of York.

The Department of Health's Policy Research Unit in Economic Evaluation of Health and Care Interventions is a 5 year programme of work that started in January 2011. The unit is led by Professor John Brazier (Director, University of Sheffield) and Professor Mark Sculpher (Deputy Director, University of York) with the aim of assisting policy makers in the Department of Health to improve the allocation of resources in health and social care.

This is an independent report commissioned and funded by the Policy Research Programme in the Department of Health. The views expressed are not necessarily those of the Department. This report should be referenced as follows:

Brazier J, Rowen D, Mukuria C, Whyte S, Keetharuth A, Rose Hole A, Tsuchiya , Shackley P. Eliciting societal preferences for burden of illness, therapeutic improvement and end of life for value based pricing: a report of the main survey Available from www.eepru.org.uk

School of Health And Related Research

THE UNIVERSITY of York

CHE
Centre For Health Economics

Acknowledgements: We are grateful to Gavin Roberts and Danny Palnoch from DH for their comments throughout this project and to Gavin Roberts for a number of important suggestions on the modelling. We would also like to thank participants at two stakeholder workshops on 13/12/11 and 28/11/12 organised by DH who contributed numerous suggestions for improving the study. However, the views expressed in this report are those of the authors alone.

EXECUTIVE SUMMARY

This report presents the findings of a study to elicit societal preferences for quality adjusted life year (QALY) gains from health care interventions. The aim was to elicit societal preferences across three characteristics: (1) burden of illness (BOI) from a medical condition given current health care interventions (i.e. QALY loss per patient from a condition due to both premature mortality measured against normal life expectancy and health related quality of life (HRQoL) below one), (2) therapeutic improvement (TI) (i.e. whether preferences for large QALY gains are disproportionately larger than the size of the gain), and (3) end of life (EOL) (defined by the National Institute for Health and Care Excellence (NICE) to be expected survival of less than 2 years and expected survival gain of 3 months or more, though survival is also examined as a continuous variable).

Methods

After undertaking a series of preparatory studies, a survey using Discrete Choice Experiments (DCE) was conducted with an online general population sample using an existing panel. Respondents were asked to choose whether to treat patient group A or B, who differed in terms of four attributes: life expectancy without treatment, survival gain from treatment, HRQoL before treatment and gain in HRQoL from treatment. These four attributes are used to derive the BOI, QALY gain and survival (or EOL). No indifference option was given. There were four different variants of the questionnaire, each having a different normal life expectancy (5, 20, 40 and 80 years) with each respondent making comparisons between groups with the same normal life expectancy.

The DCE was designed using a D-optimality algorithm which generated 580 pairs allocated across 58 blocks with each respondent completing one block, i.e. 10 DCE questions. The choices between treatment pairs A and B were analysed using conditional logistic regression with a range of specifications to estimate coefficients for the characteristics of BOI, QALY gain, and EOL. The robustness of the results across the four normal life expectancies and their sensitivity to various exclusions was examined.

Results

There were 3,669 respondents who completed the survey from an online panel. Although an age distribution similar to the UK general population was achieved, there were some differences in other characteristics. The majority spent less than 30 minutes completing the survey which comprised of: a video introduction, two practice questions (one question where one profile is dominant in QALY gains and the second question where one profile is dominant in BOI), 10 DCE tasks, nine questions on

attitudes to giving priority to patients on the basis of treatment gain, severity of illness and length of survival, and 17 questions on the health, socio-demographics and understanding of the respondent.

The practice questions indicated a strong preference for a larger QALY gain and a weak preference for favouring those with a larger BOI (all else being equal). In the practice question the proportion of all respondents choosing to treat the patients with higher BOI was 50.8%, yet following the exclusion of respondents who misunderstood the DCE task, the proportion choosing to treat the patients with higher BOI was 63.5%. There were other questions in the survey with one dominant or near dominant profile in terms of BOI (34 pairs) that also provided support for BOI. Analyses of these revealed a majority of respondents typically choosing the group with the higher BOI at 52-85%. This supports the overall finding that BOI does have a positive impact overall.

Regression results indicated that survey participants preferred to treat patients who had larger QALY gains, but at a diminishing rate and there was no support for TI. They also preferred to treat patients who had a shorter life expectancy or who were at the EOL. The results for BOI were less robust across alternative model specifications, but suggested some support for BOI. Excluding respondents who misunderstood the DCE task, regressions estimated positive, significant and robust coefficients for BOI.

The attitudinal questions seemed to broadly support these findings. They indicated limited support for giving priority to treating those who were very ill. There was widespread support for treatment gain, but a similar number of respondents wanted to give an equal priority to all patients. The support for EOL was only significant where life expectancy was below normal or they had a good health status following treatment. These attitudinal questions must be treated with caution since they do not explore respondent's willingness to trade.

Discussion

At the start of the study the hypotheses were that BOI, TI and EOL would have a positive effect on preferences. The results indicate modest support for BOI as a consideration when weighting QALYs. The evidence did not support the idea of therapeutic improvement. There was robust and consistent support for EOL in general, though responses to the attitudinal questions did not support EOL. Overall there seems to be a strong preference for larger QALY gain but at a diminishing rate. These results support the argument that the social value of a QALY is not equal between recipients, but depends on the burden of their condition and expected survival.

TABLE OF CONTENTS

EXECUTIVE SUMMARY 4

TABLE OF CONTENTS 6

List of Tables 8

1 INTRODUCTION 9

1.1 Preparatory work 11

1.1.1 Study 1: Developing a framework 11

1.1.2 Study 2: Developing and testing survey methods 13

1.1.3 Study 3: Qualitative study of protocol 15

1.1.4 Study 4: A comparison of methods and mode of administration 16

2 METHODS 19

2.1 Framework of Analysis 19

2.2 Elicitation technique 22

2.2.1 Selection of attributes and levels 22

2.2.2 Experimental design 23

2.2.3 Questionnaire design 24

2.2.4 Survey sample and scenario presentation 25

2.3 Analysis of data 28

2.3.1 Model type 28

2.3.2 Alternative absolute variables for the econometric models 29

2.3.3 Relative or standardised models 30

2.3.4 Model specification 32

2.3.5 Model performance 33

2.3.6 Robustness of results 33

3 RESULTS 34

3.1 Sample 34

3.2 Respondent views of the survey 35

3.3 Practice questions 36

3.4 Regression results 36
3.4.1 HRQoL and life expectancy before and after treatment 36
3.4.2 QALY gains and BOI 38
3.4.3 QALY gains and life expectancy before treatment or end of life 39
3.4.4 QALY gains and HRQoL and life expectancy before treatment 41
3.4.5 Comparison of model performance 42
3.4.6 Robustness of results 43
3.5 Attitudinal questions 46
3.5.1 Burden of illness (BOI) 47
3.5.2 End of life (EOL) 47
3.5.3 Therapeutic improvement (TI) 48
3.5.4 Robustness of results 49
4 DISCUSSION 49
4.1 Summary 49
4.2 QALY gains and therapeutic improvement 49
4.3 End of life 50
4.4 Burden of illness 50
4.4.1 Life expectancy, end of life and health related quality of life 51
4.5 Normal life expectancy (N) 52
4.6 Existing literature 52
4.7 Limitations 53
5 CONCLUSION 55
6 REFERENCES 56
7 Appendix 59
Appendix 1: Standardised descriptive statistics 59
Appendix 2: Results excluding respondents who misunderstood the DCE task 60
Appendix 3: Analysis on attitudinal questions and DCE responses 66

List of Figures

Figure 1 Representation of the WHO Global Burden of Disease 18

Figure 2 Representation of profile used in survey 19

Figure 3 Practice Question 1 when normal life expectancy=20, first screen 25

Figure 4 Practice Question 1 when normal life expectancy=20, second screen when respondent chose to treat patient group A 26

List of Tables

Table 1 Survey attributes and levels 20

Table 2 Summary of statistics derived from the DCE design 21

Table 3 Variables used in the models 30

Table 4 Socio-demographic characteristics 33

Table 5 Responses to practice questions 34

Table 6 HRQoL and life expectancy before and after treatment 35

Table 7 QALY and BOI 36

Table 8 QALY and split BOI 37

Table 9 QALY gains with life expectancy and HRQoL before treatment 38

Table 10 QALY gains and life expectancy before treatment/end of life 39

Table 11 QALY and BOI, and QALY and split BOI, excluding respondents who misunderstood the DCE task..... 42

Table 12 Responses to the attitudinal questions 43

1 INTRODUCTION

Economic evaluation is used to inform decisions related to setting priorities in health care and whether health care interventions should be reimbursed. A widely used method is to enumerate the cost effectiveness of an intervention in terms of the incremental cost per Quality Adjusted Life Year (QALY) and compare this to some threshold cost per QALY to reflect displaced activities.[Claxton et al, 2013] The approach is designed to improve the efficiency of health care spending and typically assumes that a QALY is worth the same regardless of who gets it. However, agencies that use cost per QALY in their decision making often take other factors into consideration explicitly or implicitly.[Devlin and Parkin, 2004] It has long been recognised that the QALY approach can incorporate a more complex algorithm than simply assuming 'a QALY is a QALY'.[Weinstein, 1988] At the same time there is emerging evidence that members of the public can weight some QALY gains more highly than others, depending on who receives them.[Green, 2009; Shah, 2009; Brazier et al, 2007; Nord, 2005; Linley and Hughes, 2013]

The literature has uncovered a broad range of attributes across which the value of QALYs may be expected to vary, including: age, health state before treatment, the size of the health benefit, socio-economic background of a typical patient, degree of responsibility, and broader notions of fairness.[Keetharuth et al, Forthcoming] For an attribute to be used in cost per weighted QALY analysis it needs to be supported by normative argument and empirical evidence to quantify its size. The empirical basis can be surveys of the general public on the grounds that they are potential tax payers or on the basis of democracy. It could also be argued that the views of politicians may be appropriate since they have a democratic mandate. Most research has been undertaken with members of the general public and this suggests that the general public does not always favour the view that all QALYs are of equal value. The main idea to emerge from the literature is that those in worse health should be given greater priority than those in better health, often referred to as the severity argument, though this is not found in all studies.[Shah, 2009]

An important consideration is the way severity is defined and measured. The earlier literature tended to focus on severity in terms of the health state of the recipient before treatment. Work by Nord and others looked at the relative weighting of an improvement in health, say from 0.2 to 0.4 on the 1-0 full health-dead scale compared to 0.8 to 1.0.[Nord, 1993; 2005] In a number of studies it was found that respondents often gave more weight to gains at the lower end of the scale. However, this is quite a narrow notion of the severity of a condition, or more generally of someone's health profile. The severity of a condition is typically seen in terms of mortality as well as the quality of their state of health. As argued by Hansson and colleagues [1994] "Severity of disease can be defined as prognosis without treatment, i.e. expected remaining life years adjusted for the quality of life for these years''.[page 353]

They went on to argue: “This implies that the same metric (such as Quality Adjusted Life Years or QALYs) can be used for comparisons of outcomes with and without treatment. If health benefit with treatment is measured along the axes of mortality, pain, physical, mental and social functions, so should severity of disease.”[Hansson *et al*.1994, page 353] It could be argued further that equity considerations should be incorporated over a person’s whole health profile and not simply starting from today. This is the basis of a fair innings criterion, whereby the weight of a QALY for a given recipient depends on what has gone before and what will happen without treatment compared to some expectation or target level of survival and health state over time.[Williams, 1997] However, decisions are made for the future and it could be argued that prospective health should be the focus for decision making.[Nord, 2005]

The research reported in this paper attempts to operationalise the notion of severity of disease set out by Hansson and colleagues[1994] using a metric that is compatible with the cost per QALY analysis used by National Institute of Health and Care Excellence (NICE).[2013] The research was commissioned by the UK Department of Health’s Medicines, Pharmacy and Industry group. In a consultation document, the English Department of Health (DH) set out a new mechanism for pricing drugs in the UK known as Value Based Pricing (VBP).[Department of Health, 2010] The proposed new mechanism will assess the cost-effectiveness of medicines by taking into account a wider scope of value, including the severity of disease and wider societal benefits.[Department of Health, 2010] The consultation specified severity in the terms set out by Hansson *et al*.[1994] and referred to this as burden of illness (BOI). BOI is measured using the outstanding QALY loss suffered by patients with current treatments.

The consultation document proposed another criterion based on the size of therapeutic improvement (TI) to reflect the benefits of those innovations that bring about a ‘step change’ in outcome for patients. This criterion implies that larger sized gains per recipient should be weighed more highly than smaller gains over more recipients. This has been examined in the literature in terms of whether a given benefit of health should be concentrated for the benefit of ‘a few’ or dispersed more widely to the ‘many’. The evidence on this attribute is mixed, with respondents often favouring greater dispersion.[Olsen, 2000] However, there is some evidence to suggest that there may be a threshold below which a given QALY gain has proportionately less value. For example in Dolan *et al*.[2008] respondents tended to prefer to disperse life years until a threshold of 2.6 years, whereas in Rodriguez-miguez and Pinto Prades[2002] there was a threshold at 9.1 years. The precise threshold is likely to depend on the context, but research into concentration versus dispersion provides some support for this idea of TI.

Finally, this work was being undertaken in the UK policy environment where there already exists another attribute that is used in health technology assessment (HTA) appraisals. This is the ‘end of life’ (EOL) supplementary criterion used by NICE which stipulates that a greater weight can be given to QALY gains where the recipients have a life expectancy of less than two years and a survival gain of 3 months

or more (provided the condition is a 'rare' disease).[NICE, 2009] Therefore this attribute has been included in the current research.

The aim of the research presented in this report was to inform the development of VBP in the UK by eliciting societal preferences for the following attributes of interest to DH: 1) BOI given current health care interventions – defined as the QALY loss per patient from a condition due to premature mortality and morbidity measured against normal life expectancy and reduced health related quality of life, 2) TI - whether preferences for large QALY gains are disproportionately larger than the size of the gain, and 3) EOL - defined by NICE to be expected survival of less than 2 years and a survival gain of 3 months or more,[NICE, 2009] though survival is also examined as a continuous variable. The methods developed for operationalising these attributes, the survey to elicit the preferences of the general public using a discrete choice experiment and the analysis that was undertaken are presented. The results are detailed together with a discussion of their implications for health care policy and research in the field.

1.1 Preparatory work

In order to elicit societal preferences for BOI, TI and EOL the following needed to be determined:

1) a framework that incorporates BOI, TI and EOL and excludes age of the patient group receiving the intervention,
2) a mode of administration of the survey,
3) a framing of questions for the survey, and
4) a comparison of alternative methods and modes of administration of the survey.

A series of preparatory studies were undertaken to determine 1 to 4. These are described briefly below. Readers not interested in the preparatory work should proceed to section 2.

1.1.1 Study 1: Developing a framework

The first study involved a review on the social value of a QALY to determine the existing understanding of societal preferences for a range of attributes.[Keetharuth et al, Forthcoming] The aims of the review were to: 1) review the attributes that society values in the literature, 2) provide a shortlist of attributes for consideration in the implementation of value based pricing, 3) determine whether any off-the-shelf weights exist that could be used, and 4) identify gaps in the literature as a basis for future research to enable the successful implementation of value based pricing.

Methods

First, a general search using Medline and Econlit was conducted. Second, a textbook was identified with a list of attributes used in the social value of a QALY literature. Third, separate searches were conducted using Medline and Econlit using the attributes found previously and their variations as search terms. Fourth, an update was carried out with similar terms to those used in the published reviews. Finally, key papers known to the authors were included and citation searches of these papers were undertaken.

Results

Fifteen attributes were identified in the literature. Criteria were used to determine a shortlist of these attributes applicable to value based pricing. The main problem was the lack of good quality empirical evidence on societal values. Severity was the only shortlisted attribute, yet if more evidence could be gathered, an alternative medium list consists of the following attributes: concentration versus dispersion of benefits, rarity of conditions, end of life and iatrogenic diseases. No off-the-shelf weights exist that are appropriate for implementation in VBP.

Discussion

It was concluded that since no off-the-shelf weights exist that can be used for BOI, TI and EOL for implementation in VBP, research needs to be conducted to elicit societal preferences for these attributes. Furthermore, no framework exists that enables the simultaneous consideration of BOI, TI and EOL in a societal preference elicitation task.

A framework was developed (explained in detail in section 2.1) that incorporated BOI, TI and EOL and excluded age of the patient group receiving the intervention. The framework was based upon the Global Burden of Disease (GBD) developed for the World Health Organisation (WHO) and the World Bank using a measure called the Disability Adjusted Life Year (DALY).[Murray et al, 1996] The framework is a modified version of the GBD based around the notion of QALY loss associated with a condition from the point of being considered for the introduction of the new technology (known by NICE as the decision point). This methodology is consistent with existing HTA methodology used by NICE as it is based on UK mortality and morbidity data using EQ-5D.

A research agenda was agreed with Department of Health to develop a preparatory survey based on the proposed framework to elicit preferences for BOI, TI and EOL.

1.1.2 Study 2: Developing and testing survey methods

The second study involved a large preparatory survey to pilot questions designed to elicit preferences for BOI, QALY gain and EOL.[Rowen et al, Forthcoming *a*] The questions were developed using the framework developed in study 1 and built upon the methods used in similar surveys undertaken to elicit societal preferences for a QALY.[Baker et al. 2010; Lanscar et al. 2011] The aims of the study were to: 1) examine the preference elicitation method most suitable for eliciting preferences for BOI, QALY gain and EOL, and 2) examine the suitability of the proposed selected design for each elicitation method.

Face-to-face interviews have been most widely used to elicit preferences, but recently there has been increased interest in collecting preference data online. Face-to-face interviews are expensive and time consuming and as a result often have lower sample sizes but provide high quality reliable data with high completion rates. Online surveys enable a large number of responses to be collected quickly and relatively cheaply but there is a concern about the quality of the data, as without an interviewer present there is no guarantee than the respondent has understood or engaged with the survey. This study used an online survey for eliciting preferences for BOI, QALY gain and EOL.

Methods

A self-complete computer-based survey was developed involving discrete choice experiment (DCE) and person trade off (PTO) tasks. Each task involves 2 groups of patients with 2 different conditions, where each group has the same normal life expectancy and HRQoL (life expectancy and HRQoL without the condition), a different lower life expectancy and/or HRQoL if they have the condition, and a different gain from treatment in terms of life expectancy and/or HRQoL. A DCE task asks whether the respondent would rather treat 100 patients with condition A or 100 patients with condition B. A PTO task determines how many patients with condition A need to be treated to be equivalent in social value to treating X patients in condition B, using an iterative procedure. Prior to launching the survey in an online environment a small pilot study was conducted of 20 respondents using the online interface in an interviewer setting. Subsequently an online survey of 3,000 respondents was conducted using a representative sample of the UK general population. At the start of the survey respondents watched a video introducing the survey and explaining the concepts involved in their survey. All respondents answered a practice DCE question and then respondents answered either a block of 10 DCE questions or 2 DCE questions followed by 5 PTO questions. Respondents were randomly allocated to a block of questions at the start of the survey. Respondents answering only DCE questions saw questions involving 2 different normal life expectancies; that is, life expectancy without having the condition. Respondents answering both DCE and PTO questions only saw one normal life expectancy. Normal life expectancy was always the same for both profiles in a task.

Each question involved a choice of which group to treat out of two groups with different conditions, Condition A and Condition B, and the question included both a text explanation of the scenarios and a diagram. Respondents were told that there were 100 patients in each group in the DCE questions. In the PTO questions there were initially 100 patients in each group. First, respondents were asked which condition they preferred to treat, and then the number of patients in that group was reduced by 10 patients. This process was repeated until the respondent was either indifferent between each condition or switched between treating condition A and condition B. Then the number of people in the condition they switched from was increased by 5 patients and the question asked one final time. At the end of the survey all respondents were asked 12 questions on their health and sociodemographics.

The scenarios for the DCE questions involved scenarios specifically selected to answer the following:

- Do respondents prefer to treat the group with the larger QALY gain (where everything else is equal across the groups)?
- Do respondents prefer to treat the group with QALY gains in HRQoL or life expectancy (where everything else is equal across the groups)?
    - Is this affected by whether life expectancy is low, middle or high and whether HRQoL is low, middle or high?
- Do respondents treat the group with the largest BOI (where everything else is equal across the groups)?
    - Is this affected by whether BOI is in HRQoL, life expectancy or both?

It was hypothesised that respondents would prefer to treat the group with the larger QALY gain, be indifferent between gains in HRQoL and life expectancy, prefer to treat the group with the largest BOI and prefer to treat the group that were categorised as EOL.

In addition the survey also included a full study design when patients in each condition have 20 years normal life expectancy for both the DCE and PTO tasks and a full study design when patients have 5 years normal life expectancy using the DCE elicitation task. This enabled the estimation of regression models for each of the normal life expectancies and models that combine DCE data for normal life expectancy of 5 years and 20 years. This approach was undertaken to determine whether regression models combining data across different normal life expectancies was feasible and appropriate and to compare the results for different normal life expectancies.

Results

The results were as hypothesised regarding overall QALY gain, EOL and QALY gain separated by HRQoL and life expectancy, as respondents were more likely to treat the group with the larger QALY gain, were more likely to treat the group that were categorised as EOL, and there was no distinguishable pattern for choosing gains in health or life expectancy. However, the results did not show that BOI was important, in fact it suggested the reverse, where respondents were more likely to choose to treat conditions with lower BOI. The results also questioned whether increasing the size of the QALY gain was increasingly important to individuals.

Discussion

Both DCE and PTO tasks were feasible. The full study design for DCE for normal life expectancies of 5 and 20 years were appropriate but the coefficients in the regression models differed for the different normal life expectancies. Study design of the PTO tasks for a normal life expectancy of 5 years proved more difficult and the data provided multiple complex challenges for analysis due to the scaling of the PTO variable generated by the study design. For this reason the DCE task was selected as more appropriate.

The finding that BOI was not important was contrary to the hypothesised results. The survey design did not enable the result to be examined further in order to ascertain whether the finding correctly reflected respondents' societal values, or whether respondents misunderstood the concept or the question (for example respondents may have chosen which condition they preferred to live in or thought was best, rather than which condition they would rather treat). The explanation for the finding could not be determined using this survey data alone as no data were collected on respondents' underlying attitudes, or using different framing of questions or a different mode of administration. There was a concern that the framing of questions and online setting may have influenced the results.

1.1.3 Study 3: Qualitative study of protocol

The third study undertook a qualitative survey to further explain the findings of study 2.[Rowen et al, Forthcoming *a*] The aim of this study was to provide information about whether the finding in the large preparatory survey undertaken in study 2, that BOI was not important, was due to genuine preferences or a misunderstanding of the questions.

Methods

The study involved qualitative interviews with a small convenience sample of 21 people to examine preferences regarding BOI in detail. Respondents completed 8 of the DCE assumption testing questions

included in the preparatory survey that asked mainly whether respondents preferred to treat the group with the higher BOI. The same online interface as the preparatory survey was used but the interviewer was also present in the room. After respondents had answered the DCE questions they were then asked open-ended questions about the reasoning behind their responses. The interviewer also explained the implications of their choice to the respondent and then asked the respondent whether this information meant they wanted to change their choice. A subsample of respondents answered questions that did not include the pictures included in the preparatory survey to determine whether this affected understanding and choices.

Results and Discussion

The results from this survey suggested that the finding in the preparatory survey that BOI was not important may at least in part be explained by a misunderstanding of the concept and the task. The results implied that the framing of the questions was not appropriate as some respondents misunderstood the questions. It also raised concerns about the suitability of the online mode of administration for this type of survey where there is little means of clarifying the questions to respondents or ensuring that they have understood the meaning of their choice. The results of this study suggested that further research was needed to develop and test the framing of questions and to determine whether an online survey is appropriate for eliciting preferences for BOI, QALY gain and EOL, or whether a face-to-face interview is more appropriate where interviewers can repeat explanations and monitor understanding and engagement with the survey.

1.1.4 Study 4: A comparison of methods and mode of administration

The fourth study examined different framings of questions and mode of administration to examine which was most appropriate for eliciting preferences for BOI, QALY gain and EOL using the framework developed in study 2.[Rowen et al, Forthcoming *b*] This was due to concerns raised in study 3 about the suitability of an online mode of administration and the framing of questions used. The aim of this study was to determine the most appropriate mode of administration and framing of questions suitable for eliciting societal preferences for BOI, QALY gain and EOL.

Studies 2 and 3 raised concerns about the elicitation of preference data online, and study 3 raised concerns about the exact wording and framing effects of questions used to elicit preferences for BOI, QALY gain and EOL. There are concerns about the quality of data from online surveys, where respondents' understanding is not easily monitored, yet face-to-face interviews are expensive and time consuming for surveys with large sample sizes. Therefore this study compared data collected online and face-to-face.

Questionnaire wording is of extreme importance as responses can be affected by the exact wording used. Framing effects are well acknowledged in the economic literature but are largely unexplored in the elicitation of societal preferences. With a complex topic for the survey, it is important to ensure that questions are worded clearly and simply to enable respondents to understand and engage with the questions. This study examined different ways of wording the questionnaire.

Methods

A 6-arm survey was conducted, where 3 arms were administered in an online survey and 3 arms were administered using face-to-face interviews. DCE was selected as the preference elicitation technique using the results of study 2. Three different framings of questions were used, and each were administered using both face-to-face interviews and online. The framing differed by: the inclusion or exclusion of pictures to accompany the descriptions of the alternatives; stating the levels of the attributes with and without treatment or stating the change in the attributes from the normal level to no treatment and from no treatment to treatment; and the labelling of the no treatment, treatment and expected health profiles; and by the labelling of the groups.

Each arm used the same questions, only the framing and mode of administration differed. The questions were selected to examine whether respondents choose to treat the group with the highest treatment gain and BOI, whether respondents systematically choose to treat the group with treatment gains in HRQoL or life expectancy, and whether responses were affected by the exact attributes in the questions. Respondents were asked which group they would rather treat. The format of the questions was simplified compared to the preparatory study, study 2, and included feedback after each practice question to assist in understanding the task.

At the start of the survey respondents either watched a short video explaining the questions or in one of the arms were read a short introduction by the interviewer. Each arm had 2 practice questions that involved a dominant choice, where only treatment gain differed across the alternatives in the first practice question, and only BOI differed across the alternatives in the second practice question. Each practice question involved a detailed explanation and after making their choice respondents were given an explanation of their choice, and asked whether they still wished to treat that patient group. If respondents changed their mind, they then started the question again. This process was repeated once in the interviewer arms and was repeated up to 7 times in total in the online survey. This process was used to ensure that respondents understood the task and understood the choice they made.

Following the 2 practice questions in the interviews, respondents completed 7 interviewer-administered DCE questions, 9 self-complete questions on attitudes, 11 self-complete questions on EQ-5D and socio-demographics and then 5 or 6 (depending on the arm) interviewer-administered questions on their understanding and what they thought of the survey. The online survey followed the same process but all questions were self-complete. At the end of the interview the interviewer reported their perception of respondent understanding, effort and concentration.

Responses in all arms were compared to determine whether responses differed by arm to the DCE questions, whether responses to the DCE questions were consistent with responses to the general attitudinal questions and what respondents thought of the survey.

Results

In the first practice question, respondents overwhelmingly chose to treat the group with the highest treatment gain, varying from 78.3% to 95.7% across arms. In the second practice question results for BOI were mixed, where 39.1% to 60.9% chose to treat the group with the higher burden across arms. A larger number of respondents in the online arms changed their mind in both of the practice questions when provided with an initial explanation of their choice (6.2% and 6.0% for the first and second practice questions respectively for respondents in the online arms versus 0% and 4.3% for respondents in the interviewer arms).

Responses to the BOI questions demonstrated that respondents did not always choose to treat the group with the higher BOI. Respondents did not always choose to treat the group with gains in either HRQoL or life expectancy. The results for BOI and treating the group with gains in either HRQoL or life expectancy differed by arm, but there was no clear pattern in differences in responses across different modes of administration or framing.

Responses to the attitudinal questions indicated that the majority of respondents believed that the NHS should give the same priority to treating all patients who are ill, regardless of how ill they are. There was limited support for BOI, no or little support for EOL, and some support for treating the group with the higher treatment gain. Overall, responses to the attitudinal questions did not differ across the arms.

Discussion

The mode of administration and framing of questions used in each survey arm was feasible, and respondents reported that they found the DCE and attitudinal questions easy to understand. In terms of understanding and wording of the questions, there was little to choose between the arms as all modes

of administration and framing of questions seemed feasible and appropriate. Respondents reported that the video was useful and pictures were helpful in the arms where they were included. The revised format of the practice questions which involved additional information and an explanation of the respondents implied choice increased the number of respondents making the expected choice.

For the final survey the mode of administration of an online survey was chosen over an interviewer survey due to lower cost and quicker data collection. The selected framing of questions contained text with pictures since respondents found them helpful and performed similarly to other arms.

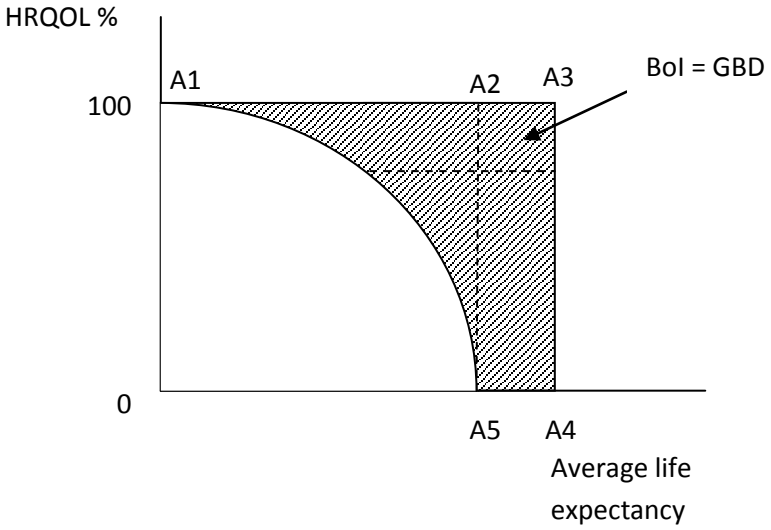
2 METHODS

The preparatory studies described above determined the framework of analysis, the elicitation technique, and mode of administration and framing of questions. In this section, these aspects of the main survey are described in more detail.

2.1 Framework of Analysis

BOI is measured using the outstanding QALY loss suffered by patients with current treatments. The same idea underpins GBD developed for the WHO and the World Bank using DALY measure.[Murray et al. 1996] The GBD study measured BOI in terms of loss of life expectancy from premature mortality and disability. BOI is represented in Figure 1 as two components: 1) the loss from premature death represented by the area A2.A3.A4.A5 and 2) the loss from disability which is the shaded area A1.A2.A5 between the profile and the line indicating full health.

Figure 1: Representation of the WHO Global Burden of Disease

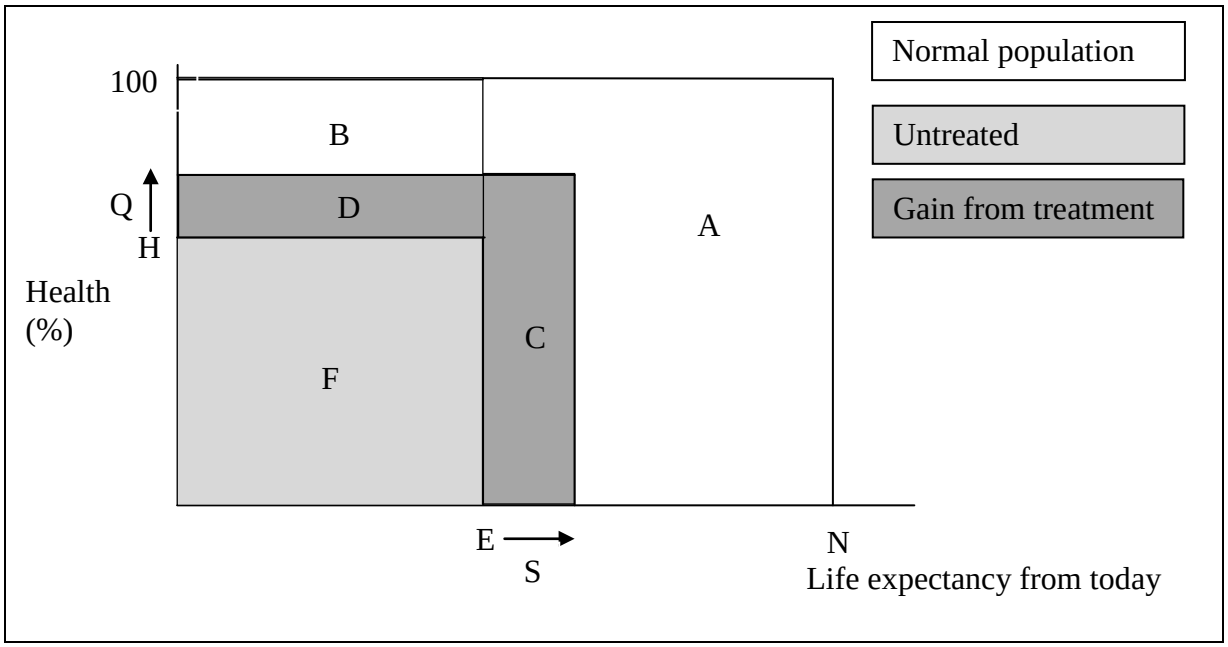


The GBD is widely used to indicate the burden of a disease and has been used for many years to inform international health policy. The most recent update moved from using expert opinion to general public surveys to provide disability weights as well as removed age and time preference weighting which had been used in previous estimates.[Salomon et al. 2012] Furthermore, expert groups were used to describe conditions in terms of their 'major functional consequences', such as the effect on daily activities and pain rather than as disability classes for valuation purposes. This brings the disability weights more in line with approaches that are used to generate weights for HRQoL measures. However, it is based on an international model life table to estimate life expectancy, whereas for UK policy it is preferable to use UK mortality data life tables. Second, the disability weights for each condition are based on the impact of a given disease rather than using an empirically based method with a common measure of HRQoL such as the EQ-5D which is used in HTA in the UK. The measurement of BOI used in this study is related to GBD at the level of an individual, except that it is measured from the point at which the treatment decision is being considered, such as patients with rheumatoid arthritis who have not responded to first line treatment and are being considered for second line treatment. It does not consider a person's previous health.

The proposed methodology is more in keeping with existing HTA methodology currently used in the UK by NICE as it can be based on UK mortality and morbidity data using EQ-5D. The components of the attributes of BOI, TI and EOL as described in this study are shown in Figure 2. Individuals in this population have a health profile without the new treatment, which for simplicity is represented by health state (H) and life expectancy (E), and this is assumed to be without treatment. To estimate the BOI or QALY loss associated with the condition it is necessary to establish an expected or target level of health and life expectancy. The target level of health without the disease is assumed to be 100% with

normal life expectancy (N). The expected improvement from treatment is represented by a gain in HRQoL (Q) and an improvement in survival (S).

Figure 2: Representation of profile used in survey



Where N = normal life expectancy, E = life expectancy without treatment, S = survival gain from treatment, H = health before treatment, Q = health gain from treatment

The information presented in Figure 2 includes:

- Normal health without the condition - HRQoL and life expectancy without the condition (100%*N or 1*N where 1 represents full health in utility terms)
- A+C represents the burden of illness from a shorter life expectancy without treatment i.e. QALY loss from premature mortality (BOISU)
- B+D represents the burden of illness from a poorer HRQoL without treatment i.e. QALY loss from morbidity (BOIQL)
- A+B+C+D represents burden of illness i.e. effect of having the condition without treatment (BOI)
- C represents QALY gain as a result of gains in life expectancy from treatment (QLYSU). Note that the area represented by Q*S is included in these gains from life expectancy rather than those from HRQoL as these QALY gains are not possible without extending survival.
- D represents QALY gain as a result of gains in HRQoL from treatment (QLYQL)
- C+D represents overall QALY gains from treatment
- F represents HRQoL and life expectancy without treatment (equivalent to N- BOI)
- EOL is represented by life expectancy before treatment i.e. E less than or equal to 2 years.

2.2 Elicitation technique

A DCE based on pairwise comparisons was chosen as the method to elicit preferences. DCE was chosen since it permits the simultaneous consideration of different attributes in a format that is amenable to being administered online. It had also already been successfully employed by Baker and colleagues in their social value of a QALY project.[Baker et al, 2010] Each pairwise task asked respondents to choose which group they would prefer to treat, patient group A or patient group B. No indifference option was given. The profile of normal health without the condition was the same across both patient groups, but their health and life expectancy both with and without treatment could vary (Figure 3 shows an example using the first practice question).

2.2.1 Selection of attributes and levels

The profiles of patient groups A and B in each pairwise comparison were generated using 4 attributes: life expectancy without treatment (E), survival gain from treatment (S), health before treatment (H), health gain from treatment (Q) (see Figure 2). The levels of each of the attributes for different levels of normal life expectancy are outlined in Table 1. These were selected to cover a full range of potential levels, but ensure precision over the more common interventions in the real world. Some interventions are for patient groups involving mainly children and it is possible that such interventions could have large QALY gains. However most HTAs conducted in the UK involve interventions used in patients with a small number of years of normal life expectancy left with small QALY gains. For example, Walker[2011] found that median QALY gains submitted to the Scottish Medicines Consortium (SMC) were 0.097, and only 1 intervention in 8 had mean QALY gains greater than 1 QALY. Therefore, additional levels are included when normal life expectancy is 5 years (i.e. patients are at the end of their natural life) to obtain greater precision for interventions used at the end of life and to obtain precision for the types of HTAs often conducted in the UK. The levels of each of the attributes were determined following input from the DH.

Table 1: Survey attributes and levels

Attribute	**Levels**	**Levels**	**Levels**	**Levels**
Normal life expectancy, N	5 years	20 years	40 years	80 years
Life expectancy without treatment, E	3 months 6 months 9 months 1 year 2 years 5 years	3 months 1 year 2 years 5 years 10 years	3 months 1 year 2 years 5 years 10 years 30 years	3 months 1 year 2 years 5 years 10 years 30 years 60 years
Survival gain from treatment, S	0 1 month 3 months 6 months 9 months 1 year 3 years	0 3 months 6 months 1 year 3 years 10 years	0 3 months 6 months 1 year 3 years 10 years	0 3 months 6 months 1 year 3 years 10 years 60 years
Health before treatment (%), H	10 20 40 60 80	10 20 40 60 80	10 20 40 60 80	10 20 40 60 80
Health gain from treatment (%), Q	0 2 5 10 20 30 60	0 2 5 10 30 60	0 2 5 10 30 60	0 2 5 10 30 60

2.2.2 Experimental design

A full factorial design using the attributes and levels specified in Table 1 would result in a large number of possible profiles, meaning it is infeasible to conduct a valuation survey involving every possible profile. Profiles were selected using a D-optimality algorithm [Carlsson and Martinsson, 2003; Kuhfeld, 2005] and the true model was specified in such a way as to allow for the estimation of all parameters of interest. Impossible profiles (such as profiles involving a HRQoL gain of more than 100%) were excluded from the candidate set for the design. In total the DCE designs constituted 580 pairs of profiles, with the number of pairs varying across designs depending on the number of attributes and levels in the design. Pairs were allocated into 58 combinations (also known as 'blocks') of 10 pairs. Each combination contained pairs for one normal life expectancy. Summary statistics of the attribute combinations generated by the DCE design are reported in Table 2 for each level of life expectancy. These show the large range of QALY gains considered in this survey starting from 0.005 up to 63, and BOI from 1 to 80 QALYs lost and life expectancy from 0.25 to 60 years.

Table 2: Summary of statistics derived from the DCE design

Normal life expectancy		**5**	**20**	**40**	**80**
DCE pairs		160	120	140	160
QALY gain	Mean (s.d.)	0.673 (0.744)	2.039 (2.638)	3.012 (4.219)	8.510 (13.64)
	Median	0.4	0.9	1.2	2.45
	Minimum	0.005	0.0125	0.005	0.005
	Maximum	3.4	16	21	63
QALY gain due to survival	Mean (s.d.)	0.442 (0.642)	1.417 (2.336)	1.462 (2.389)	5.792 (12.722)
(QLYSU)	Median	0.2	0.4	0.4	0.6
	Minimum	0	0	0	0
	Maximum	3	10	10	60
QALY gain due to health	Mean (s.d.)	0.231 (0.427)	0.623 (1.205)	1.549 (3.490)	2.718 (6.416)
(QLYQL)	Median	0.075	0.1	0.2	0.3
	Minimum	0	0	0	0
	Maximum	3	6	18	36
BOI	Mean (s.d.)	4.507 (0.651)	18.518 (1.931)	36.487 (5.449)	73.349 (10.899)
	Median	4.7	19.2	39	78.4
	Minimum	1	12	16	32
	Maximum	4.975	19.975	39.975	79.975
BOI from premature mortality	Mean (s.d.)	3.795 (1.23)	16.379 (3.585)	31.335 (10.696)	64.401 (20.582)
(BOISU)	Median	4.25	18	35	75
	Minimum	0	10	10	20
	Maximum	4.75	19.75	39.75	79.75
BOI from poor health status	Mean (s.d.)	0.712 (0.852)	2.139 (2.482)	5.153 (7.312)	8.948 (13.421)
(BOIQL)	Median	0.4	1	1.8	3
	Minimum	0.05	0.05	0.05	0.05
	Maximum	4.5	9	27	54
Life expectancy untreated	Mean (s.d.)	1.206 (1.23)	3.621 (3.585)	8.665 (10.696)	15.599 (20.582)
(E)	Median	0.75	2	5	5
	Minimum	0.25	0.25	0.25	0.25
	Maximum	5	10	30	60

2.2.3 Questionnaire design

Selected attributes and levels were used to design the DCE profiles in terms of life expectancy and health before and after treatment. Each scenario in the survey had two profiles representing condition A and B with information presented in text as well as using a graph. An example practice question is shown in Figure 3. Respondents were asked to imagine there are the same number of patients in two different groups, patient group A and patient group B, each with a different medical condition, where each medical condition can affect the health of the patients and how long they live. They were told that

patients can be treated, but it is only possible to treat one patient group as there are not enough resources such as money, staff and hospital space to treat all patients. Respondents were then asked which patient group they thought the NHS should treat. Profile information for each patient group was presented alongside each other using the following information:

- Number of years lived from today without treatment - E
- HRQoL (in %) without treatment - H
- Number of years lived from today with treatment – E+S
- HRQoL (in %) with treatment – H+Q.

Participants were not presented with S and Q in the text, though these are easily calculated. There was additional information on the level of health (100%) and the number of years that patients would live without the medical condition (N). This was a single line of text covering both profiles that was fixed in each scenario but varied across questionnaire variants where normal life expectancy changed. There were 4 different variants of the questionnaire; each had a different level of normal life expectancy - 5 years, 20 years, 40 years or 80 years. Normal life expectancy was fixed across all pairs for each participant. This was due to concerns that different normal life expectancies would be confusing for respondents and would highlight the differences in age of the two profiles, where age is a consideration that is not regarded as politically desirable in the UK NHS.[Department of Health, 2012] Each profile also had a graph (Figure 3) which provided BOI and QALY information in QALY terms with different colours highlighting expected QALYs without the condition, expected QALYs without treatment, and QALY gains from treatment.

2.2.4 Survey sample and scenario presentation

Respondents from an online panel were contacted via email to participate in the survey. Respondents were sampled to be representative of the UK in terms of age and gender. At the start of the survey respondents read an information sheet and gave informed consent to participate in the survey. Respondents were then shown a short video explaining the questions including descriptions of full health and less than full health using dimensions and levels of health found in the EQ-5D and the graph format illustrated in Figure 3. It could not be guaranteed that respondents watched the video, but the video had to be played in full before the respondent could proceed to the practice questions.

The survey had 2 practice questions, each of which involved an explanation of their choice with an opportunity for respondents to change their mind multiple times. The first practice question was dominant in QALY gains in one profile while all other attributes were the same. The second was dominant in BOI in one profile while all other attributes were the same. Figure 3 shows the information

displayed on the first screen of practice question 1 with a normal life expectancy of 20 years. Practice questions highlighted what individuals should focus on, including the life of each patient group without treatment, with treatment and without the medical condition (normal life expectancy and health) in an additional box below the scenario (Note: this box was not present in the actual survey scenarios). Respondents were also informed that only 1 patient group could be treated. Figure 4 shows the explanation that was displayed on the second screen for a respondent who chose patient group A in practice question 1. Respondents were asked on the second screen whether they still wished to treat that group, and started the next question if they did not change their mind, or were shown the question again from the first screen if they did change their mind. The same format was used if the respondent chose patient group B. Respondents were allowed up to 7 attempts at each practice question before moving on automatically to the next question.

After the 2 practice questions, respondents self-completed 10 DCE questions, 9 questions on attitudes, and 17 questions covering EQ-5D, socio-demographics, understanding and what they thought of the survey. In each DCE question, one or more of the elements presented in the scenario varied. Normal life expectancy and normal HRQoL did not vary for individual participants.

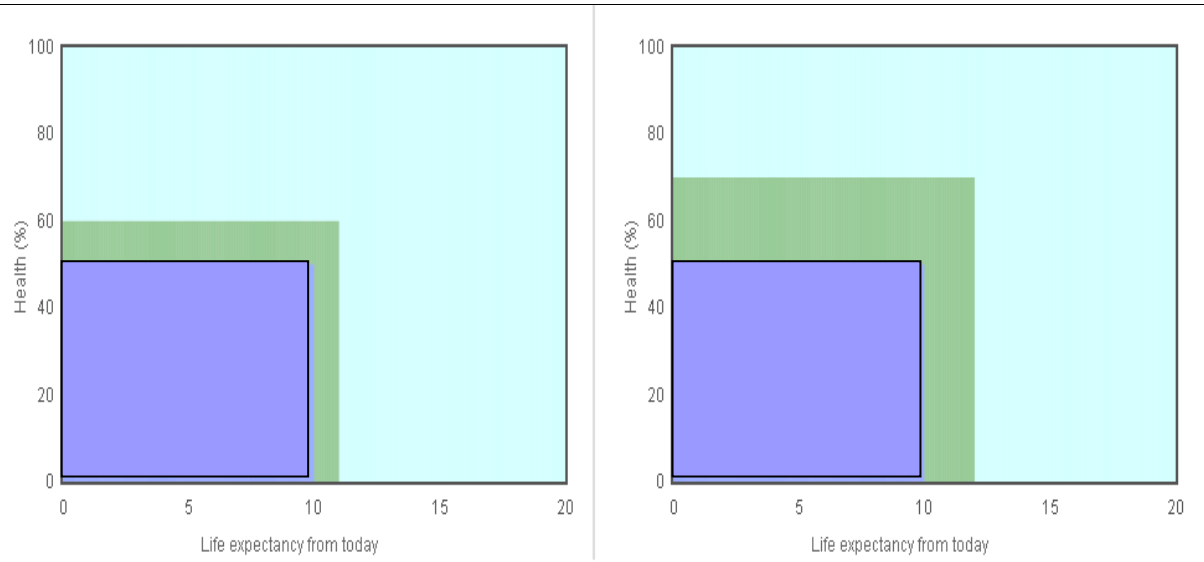
Attitudinal questions were included to determine respondents' general views on BOI, TI and EOL. This enables further interpretation of the results of the practice questions and DCE analysis, as these should be in accordance with the results of the attitudinal questions that remove the complexities and intricacies of the DCE questions. The attitudinal questions were selected and worded with input from the DH. The questions were worded as simply as possible to try and ensure that they were not misunderstood, but this means that they involved a simplification of BOI, TI and EOL.

Figure 3: Practice Question 1 when normal life expectancy=20, first screen

Both groups of patients have a medical condition, and this affects their health and how long they live

Patient group A	**Patient group B**
Without treatment • will live for **10** years from today • with **50%** health	**Without treatment** • will live for **10** years from today • with **50%** health
With treatment • will live for **11** years from today • with **60%** health	**With treatment** • will live for **12** years from today • with **70%** health

If these patients did not have a medical condition
they would live in 100% health for 20 years (**N**) from today



Without treatment Treatment gain No condition

Please make sure you consider in your answer:
- the life of each patient group without treatment
- the life of each patient group with treatment
- the life of each patient group if they did not have a medical condition

There are the same number of patients in each patient group.

Remember that you can treat only 1 patient group.

The patient group you do not treat will live the life without treatment.

Only 1 patient group can be **treated**, the other patient group will live for the rest of their life **without treatment**
Which patient group do you think the NHS should treat?

Patient group A	Patient group B
□	□

Figure 4: Practice Question 1 when normal life expectancy=20, second screen when respondent chose to treat patient group A

Both groups of patients have a medical condition, and this affects their health and how long they live	
Patient group A **Without treatment** • will live for **10** years from today • with **50%** health	**Patient group B** **Without treatment** • will live for **10** years from today • with **50%** health
With treatment • will live for **11** years from today • with **60%** health	**With treatment** • will live for **12** years from today • with **70%** health

The impact on how long the patients live and their health from having the medical condition was the same for both patient groups.

You chose that the NHS should treat patient group A.

These patients will live for 11 years from today with 60% health.

Patient group B will not be treated. These patients will live for 10 years from today with 50% health.

You have chosen the treatment that gives the smallest treatment gain.

Do you still think that the NHS should treat patient group A?

Yes □ No □

2.3 Analysis of data

2.3.1 Model type

The DCE data was modelled based on a random utility (RUT) framework.[Luce, 1959; McFadden, 1974] Within the RUT framework, utility U_{ij} for an individual *i* is assumed to be a function of an explainable utility component V_{ij} and a random component ε_{ij}:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \quad (1)$$

where *j* represents the alternatives individuals have within a choice set. The alternative chosen by the individual is assumed to confer greater utility than the other alternative. Choices are based on a

set of attributes captured in V_{ij}. Other influencing factors that are not observed are captured by the random component. DCE data provide the alternatives that individuals have chosen, in this case whether respondents were willing to treat patient group A or patient group B and these are modelled using the conditional logistic model which models the probability that individual *i* chooses profile *j*, so that the probability of an individual choosing to treat patient group A over patient group B (P_A) is given by:

$$P_A = \frac{\exp(V_A)}{\exp(V_A) + \exp(V_B)} \quad (2)$$

where V_A and V_B represent the utility that the person derives from choosing to treat patient group A and B, respectively. V, is modelled as a function of a vector of attributes z.

$$V = f\,(\mathbf{z}) \quad (3)$$

The vector z contains the variables from the DCE. Participants in the survey used the information in the questionnaire text, the graph or both (Figure 3) to make decisions about which profile was preferred. This may have been based on absolute values, where the concern is the difference between health before and after treatment or relative differences that are standardised in some way. Variables are described first in terms of absolute values and second in terms of relative variables.

2.3.2 Alternative absolute variables for the econometric models

1) Life expectancy and HRQoL before (H, E) and after treatment (H+Q, E+S)

The simplest model that is closest to the way data are presented to respondents is in the form of life expectancy and HRQoL before and after treatment. Life expectancy and HRQoL before treatment are E and H respectively and are presented in the text describing the alternatives. Life expectancy and HRQoL after treatment are these variables summed with the changes in life expectancy and HRQoL of S and Q respectively. E and H are expected to be negative and E+S and H+Q are expected to be positive if respondents give greater weight to those in poorer health and those getting larger gains respectively.

2) QALY gains from treatment and burden of illness from condition (BOI)

This model is more relevant from a policy perspective, since one QALY gain is the standard numeraire used in cost effectiveness assessments of new technologies by NICE,[2013] SMC,[2013] and AWMSG.[2012] QALY squared was included to capture the effect of TI. BOI is defined as it is in Figure 2 as being composed of QALYs lost from a condition compared to a normal life expectancy and 100% HRQoL. It is composed of losses from premature mortality and morbidity. All the coefficients on these terms are expected to be positive if respondents give greater weight to QALY gains, TI (QALY squared term) and BOI.

A related model examines splitting BOI into losses from premature mortality (BOISU) and morbidity (BOIQL). Coefficients on these terms are also expected to be positive.

3) QALY gains from treatment and end of Life (EOL)

This model examines the role of life expectancy before treatment or pre-treatment survival in respondent choices. Life expectancy (E) is included separately as a continuous variable in one model or as a dummy variable representing EOL (i.e. E of 2 years or less with survival gains of 3 months or more representing NICE criteria for end of life)[NICE, 2009] in another model. E is expected to be negative. EOL is expected to be positive if respondents give greater weight to shorter life expectancy with current treatment but where new treatment offers gains of 3 months or more.

It should be noted that BOI and EOL are not entered into the same model given the conceptual overlap between the two. This is reflected in the way E is a key driver in BOI (BOISU and BOIQL).

It was also decided for completeness to run a model with QALY and E and H rather than BOI. E and H are expected to be negative if respondents give greater weight to those in poorer health with current treatment.

2.3.3 Relative or standardised models

Respondents' choices may be driven by the relative sizes of the QALY gain or BOI, rather than their absolute value. There is some support for this in the literature, such as the notion of proportional shortfall described by Johannesson[2001] and Stolk et al.[2005] where BOI is expressed as a proportion of the loss in QALYs from the condition divided by the QALY profile expected without the condition. The proportional shortfall idea has been incorporated into health care assessment in the Netherlands and it was decided to examine it here. The idea of standardising was extended to QALY gains and BOI used in the current research.

The original small scale study by Stolk and colleagues takes into account from the moment of intervention, expected QALYs without treatment relative to expected QALYs without the condition; i.e. if there was a 'wonder pill' that could restore the patient to full health.[Stolk et al, 2005] This would equate to baseline health i.e. health and life expectancy without treatment (area F in Figure 2) relative to overall health (1*N) in the models estimated here. Stolk et al.[2005] operationalize proportional shortfall as 1-(expected QALYs without treatment /expected QALYs without the condition) so that higher values represent conditions with higher burden. In the models estimated here, this is equivalent to 1 minus BOI divided by N; i.e. 1-BOI/N = (N-F)/N = 1-F/N. It is important to bear in mind that the current study did not have a wonder pill profile. Results could be affected by the scale of life expectancy that respondents saw (5, 20, 40 and 80 years) and standardising the different questionnaire variants enables comparisons. QALYs and terms for survival may be influenced by N, and so models have been examined where they are standardised in the same way by dividing by N. All models where variables are divided by N are referred to as standardised for consistency.

The variables used in the models presented in this report and how they are derived are summarised in Table 3.

Table 3: Variables used in the models

Variable	Description	Text (Figure 3)	Derived
H	HRQoL with condition and **without** treatment	H	-
E	Life expectancy with condition and **without** treatment	E	-
HQ	HRQoL with condition and **with** treatment	H+Q	
ES	Life expectancy with condition and **with** treatment	E+S	
Q	HRQoL gains from treatment	Q	-
S	Survival gains from treatment	S	-
stdH	Standardised H	-	H/100
stdE	Standardised E	-	E/N
stdHQ	Standardised HQ	-	(H+Q)/100
stdES	Standardised ES	-	(E+S)/N
N	Normal life expectancy in profile	N	-
EOL	End of Life	-	EOL=1 if E≤2 and S≥3 months
QALY	Expected quality adjusted life year gains from treatment	[C + D]	(E*Q/100)+(S((H+Q)/100))
BOI	Expected quality adjusted life year loss from condition without treatment	[A+B+C+D]= [N - F]	N – (H/100*E)
stdQALY	Standardised QALY	[C + D]/N	{(E*Q/100)+(S((H+Q)/100))}/N
stdBOI	Standardised BOI/Proportional shortfall	1-F/N= [N - F]/N	{N – (H/100*E)}/N
BOIQL	Expected quality adjusted life year loss in HRQoL from condition without treatment	[B+D]	E -((H/100)*E)
BOISU	Expected quality adjusted life year loss in survival from condition without treatment	[A+C]	N-E
stdBOIQL	Standardised BOIQL	[B+D]/N	{E -((H/100)*E)}/N
stdBOISU	Standardised BOISU	[A+C]/N	{N-E}/N

2.3.4 Model specification

The function can take many different forms. The survey was designed to estimate an additive model where each attribute is entered as an independent main effect.

$$V = \beta_{ij}\mathbf{z}_{ij} + \gamma_{ij}\,\mathbf{z}_{ij}^2 + \varepsilon \qquad (4)$$

where V represents utility, z represents a vector containing the variables described in the section above and z^2 represents squared terms. β captures the linear relationship between z and utility while γ captures non-linear effects. This model specification was chosen, as opposed to the specification used by Lancsar et al.,[2011] to keep the model as simple and transparent as possible. An additive model based on equation 4, for the simple QALY, BOI model is given by:

$$V = \beta_1 QALY + \beta_2 QALY^2 + \beta_3 BOI \quad (5)$$

The marginal rate of substitution (MRS) of BOI for QALYs is generated from equation (5) using the ratio of the marginal utility of BOI to the marginal utility of a QALY:

$$MRS = -\frac{\beta_3}{\beta_1 + 2\beta_2 QALY} \quad (6)$$

A value of 1 QALY is used in equation (6) to generate the marginal rate of substitution for illustrative purposes. The marginal rate of substitution of EOL for QALYs is generated using the equivalent regression containing terms for QALY, QALY squared and EOL.

2.3.5 Model performance

Performance of all regression models was assessed using the log-likelihood, Rho-squared, Akaike Information Criterion (AIC)[Akaike, 1973] and the Schwarz Bayesian Information Criterion (BIC).[Schwarz, 1978] Models are preferred with larger log likelihood and larger Rho-squared. Models with lower AIC and BIC are preferred, but the penalty for the inclusion of additional variables in BIC is larger than in AIC, meaning that BIC tends to favour models with fewer variables than AIC. Collinearity was also assessed using the variance inflation factor (VIF), with values greater than 10 as evidence of collinearity.

2.3.6 Robustness of results

Robustness of results was assessed by excluding responses from individuals who may have not understood or engaged with the survey. A number of exclusion criteria were examined:

- respondents who reported that they found the survey quite or very difficult
- respondents who took less than 5 minutes or more than 60 minutes to complete the survey
- respondents who selected to treat the same condition for all 10 questions (this may indicate respondents were selecting either all left or right sides of the screen)
- respondents who chose a profile that had a larger number of total lifetime QALYs after treatment, but smaller QALY gain from treatment and lower BOI before treatment than the other profile. These respondents were excluded on the grounds that they probably misunderstood the task as they chose the profile they thought was best or that they wanted to live in, not the profile with the patients who were most deserving of treatment.

- the first survey question from all respondents because it was likely to be the least reliable. Questions were allocated to respondents in a random order and therefore exclusion of the first question should have no systematic impact on results.

The regression models were re-run with each of these exclusions to examine the impact of these exclusions on the sign, significance and size of the coefficients. Finally, the consistency of the regression results with responses to the attitude questions were examined by re-estimating models for those who expressed attitudes favourable or unfavourable to BOI, TI and EOL.

3 RESULTS

3.1 Sample

3,669 respondents completed this online survey, providing a response rate of 55% of people who accessed this survey and another survey conducted simultaneously and analysed elsewhere. Sample sizes across the 4 normal life expectancy groups ranged from 760 to 1,022. Response rates were lower than in the fourth preparatory study, but this may be expected due to the more restrictive sampling undertaken in the main study which was required to achieve a target quota of respondents aged over 65 and large target numbers of respondents who are harder to obtain responses from, such as 25-44 year old males. All respondents completed every question. No respondents were excluded from the main analysis.

Characteristics of the samples are compared to the general population in England in Table 4. In comparison to the general population of England, those who completed the survey had a similar age distribution but with a higher proportion of females, unemployed individuals, long-term sick and retired, and a lower proportion of individuals who were employed or self-employed. The sample also had a lower EQ-5D score than the general population of England, indicating poorer health. Although 66.9% of individuals stated their health in general was good or very good, 37% stated that they were limited by a long term health condition or disability and 33.6% stated they had experienced serious illness in themselves. A large proportion of the sample, 48.2%, had a degree or equivalent professional qualification.

Table 4: Socio-demographic characteristics

	All respondents	England *
N	3,669	
Mean age (s.d.)	46.5 (16.6)	NA
Age distribution (years)		
18-40	39.9%	41.6%
41-65	42.1%	39.1%
Over 65	18.0%	19.3%
Female	54.3%	51.3%
Married/Partner	62.4%	NA
Employed or self-employed	47.3%	60.9%
Unemployed	6.2%	3.4%
Long-term sick	6.4%	5.3%
Full-time student	7.2%	7.3%
Retired	23.8%	13.5%
Secondary school is highest level of education	21.6%	
Degree or equivalent professional qualification	48.2%	
Health in general is very good or good	66.9%	
Limited by long term health condition or disability	37.0%	
EQ-5D score, mean (s.d.)	0.78 (0.26)	0.86 (0.23)†
Experienced serious illness in yourself	33.6%	
Experienced serious illness in family	74.5%	
Experienced serious illness in caring for others	33.5%	

Notes: * Statistics for England in the Census 2001. Questions used in this study and the census are not identical. The census includes persons aged 16 and above whereas this study only surveys persons aged 18 and above. Age distribution is here reported as the percentage of all adults aged 18 and over.
† Interviews conducted in the Measurement and Valuation of Health (MVH) study (Kind et al, 1999). Ara and Brazier (2011) report a mean EQ-5D score of 0.87 generated using the Health Survey for England. NA=Not available

The median completion times from consent to the end of the survey was 21 minutes (IQR 17-27 minutes) with the majority of respondents (≈ 80%) spending less than 30 minutes on the survey, but for a small proportion (≈ 5%) of respondents the entire survey took over an hour. Most (>80%) spent less than 10 minutes on the introduction video and practice questions, suggesting the respondent watched the video and then considered the practice questions. However, some respondents had long times (up to 1 hour) that suggest they may have left the survey idle in this time and therefore there are doubts as to whether these respondents watched the video.

3.2 Respondent views of the survey

The majority of respondents, 77.7%, reported that the DCE questions were either very easy (36.1%) or fairly easy (41.6%) to understand, varying from 76.2% to 79.5% across the different life expectancies. The majority of respondents, 77.7%, also reported that the attitudinal questions were either very easy (36.8%) or fairly easy to understand (40.9%), varying from 76.2% to 80% across the life expectancies.

3.3 Practice questions

In practice question 1, respondents overwhelmingly chose to treat the group with the highest treatment gain, patient group B, and this is consistent across the different normal life expectancy variants, varying from 90.7% to 92.5% (Table 5). In practice question 2, patient group A is the group with the higher BOI, but there is limited evidence that respondents prefer to treat this patient group with 46.8%, 54.3%, 52.3% and 50.7% across the four versions choosing this group (Table 5).

Overall 4.9% and 6.9% of respondents changed their selection following the first explanation in practice questions 1 and 2 respectively and when they were faced again with the original question 62.7% and 73.0% respectively made a different choice. Out of those, 81.2% chose to treat the group with the highest treatment gain in practice question 1, and for practice question 2, 76.0% chose to treat the group with the highest BOI. These results suggest that the explanations did clarify the profiles for some respondents and these respondents altered their choices.

Table 5: Responses to practice questions

Practice Question	Normal life expectancy	Practice question First response			Practice question Second response			Practice question Final response		
		n	A	B	n	A	B	n	A	B
1	5	1,022	7.7%	92.3%	42	26.2%	73.8%	1,022	6.8%	93.2%
	20	760	8.7%	91.3%	28	14.3%	85.7%	760	7.1%	92.9%
	40	889	9.3%	90.7%	54	22.2%	77.8%	889	7.8%	92.2%
	80	998	7.5%	92.5%	55	23.6%	76.4%	998	6.5%	93.5%
2	5	1,022	41.9%	58.1%	77	68.8%	31.2%	1,022	46.8%	53.2%
	20	760	50.4%	49.6%	45	75.6%	24.4%	760	54.3%	45.7%
	40	889	48.5%	51.5%	60	65.0%	35.0%	889	52.3%	47.7%
	80	998	47.5%	52.5%	70	61.4%	38.6%	998	50.7%	49.3%

Notes: Respondents who stated after the explanation of their choice that they did not still want to treat the same group were asked the question again, and this process was repeated allowing respondents up to 7 attempts in total before moving on to the next question.

3.4 Regression results

3.4.1 HRQoL and life expectancy before and after treatment

Table 6 summarises the results based on HRQoL before (H) and after (HQ) treatment and life expectancy before (E) and after (ES) treatment in absolute and standardised terms. As expected, H and E are negative while HQ and ES are positive across the questionnaire variants indicating that respondents were less likely to choose profiles with high levels of HRQoL and life expectancy before

treatment, and more likely to choose profiles with high levels of HRQoL and life expectancy after treatment.

Table 6: HRQoL and life expectancy before and after treatment

Variables	All	5yrs	20yrs	40yrs	80yrs
H	-0.033128***	-0.037765***	-0.035411***	-0.036326***	-0.035009***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
E	-0.087855***	-1.043415***	-0.196465***	-0.170115***	-0.035783***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HQ	0.048005***	0.053244***	0.051889***	0.048480***	0.051085***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ES	0.095849***	1.178916***	0.241505***	0.172891***	0.044745***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	-20433	-5173	-3980	-4980	-5083
Rho-squared	0.197	0.270	0.244	0.192	0.265
AIC	40875	10354	7969	9967	10174
BIC	40912	10385	7999	9999	10205
Mean VIF	2.12	2.18	1.88	6.07	1.79
stdH	-3.584132***	-3.776524***	-3.541122***	-3.632648***	-3.500938***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdE	-4.544852***	-5.217076***	-3.929310***	-6.804589***	-2.862620***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdHQ	5.088076***	5.324445***	5.188886***	4.847972***	5.108451***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdES	5.047739***	5.894578***	4.830097***	6.915660***	3.579621***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	-19327	-5173	-3980	-4980	-5083
Rho-squared	0.240	0.270	0.244	0.192	0.265
AIC	38662	10354	7969	9967	10174
BIC	38699	10385	7999	9999	10205
Mean VIF	2.12	2.18	1.88	6.07	1.79
Observations	73,380	20,440	15,200	17,780	19,960

pval in parentheses; *** p<0.01, ** p<0.05, * p<0.1; H and E HRQoL and life expectancy before treatment; HQ and ES HRQoL and life expectancy after treatment

The coefficients on HRQoL terms are stable across the questionnaire variants while life expectancy terms vary with normal life expectancy. The absolute size of the coefficients on life expectancy without treatment (E) and life expectancy with treatment (ES) decreases as normal life expectancy increases, with the largest difference between 5 and 20 years. Standardisation reduces the differences in the life expectancy terms but does not remove them completely and there is an anomaly at 40 years. These results suggest that although normal life expectancy (N) did not vary for individual respondents, it may have influenced the results.

3.4.2 QALY gains and BOI

Table 7 summarises the results based on QALY gains and BOI (QALY loss). As expected QALY gains are positive; BOI is small and mainly positive, indicating that respondents preferred profiles with higher QALY gains and also higher BOI. QALY squared is negative, indicating that QALY gains were preferred at a decreasing rate. BOI squared was tested but did not improve the models and although it was statistically significant in some models, BOI was no longer statistically significant, hence the squared term has not been included here. The marginal rate of substitution of 1 unit of BOI is -0.027 QALYs using the 80 year variant and the regressions using non-standardised variables.

Table 7: QALY and BOI

Variables	All	5yrs	20yrs	40yrs	80yrs
QALY	0.275717***	3.641225***	0.750890***	0.403637***	0.171109***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
QALY_sq	-0.003827***	-0.708627***	-0.037264***	-0.014023***	-0.002050***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
BOI	0.017239***	0.119966***	-0.000024	0.038750***	0.004505**
	(0.000)	(0.000)	(0.998)	(0.000)	(0.015)
Log likelihood	-21775	-5160	-4043	-5246	-5416
Rho-squared	0.144	0.272	0.232	0.149	0.217
AIC	43555	10326	8093	10498	10838
BIC	43582	10350	8116	10521	10861
Mean VIF	5.35	6.92	4.77	5.39	7.42
stdQALY	14.9943***	18.20613***	15.01781***	16.14547***	13.68874***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdQALY_sq	-14.80278***	-17.71567***	-14.90545***	-22.4361***	-13.11786***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdBOI	0.758085***	0.599828***	-0.000477	1.549989***	0.360379**
	(0.000)	(0.000)	(0.998)	(0.000)	(0.015)
Log likelihood	-20004	-5160	-4043	-5246	-5416
Rho –squared	0.213	0.272	0.232	0.149	0.217
AIC	40015	10326	8093	10498	10838
BIC	40042	10350	8116	10521	10861
Mean VIF	5.70	6.92	4.77	5.39	7.42
Observations	73,380	20,440	15,200	17,780	19,960

pval in parentheses; *** p<0.01, ** p<0.05, * p<0.1; QALY – QALY gains from new treatment; QALY_sq – QALY squared; BOI – QALY loss from condition with current treatment

The coefficient size of QALY gains and QALY squared falls as normal life expectancy increases with the largest difference between 5 and 20 years (Table 7). The coefficient size of BOI also falls between 5 and 80 years, but not monotonically. Standardising the QALY coefficients by dividing QALY gain by N substantially reduces the differences between the four questionnaire variants. By contrast, standardising BOI does not reduce the differences between the coefficients and so the results would not seem to support the proportional shortfall model.

Table 8 summarises the results where BOI is split between premature mortality and morbidity. QALY and QALY squared are positive and negative respectively as before and similar in magnitude to the previous models. In the pooled data set, the BOI from HRQoL loss (BOIQL) is negative while burden from life years (BOISU) lost is positive. This indicates that respondents were more likely to choose profiles with higher burden from life years but lower burden in HRQoL. These terms are not consistent in direction between the four questionnaire variants, though these coefficients are not significant across all the questionnaires.

Standardisation substantially reduces differences in the coefficients on the QALY terms between the four variants, but large differences remain in the BOIQL and BOISU coefficients.

Table 8: QALY and split BOI

Variables	All	5yrs	20yrs	40yrs	80yrs
QALY	0.309185***	3.626376***	0.783542***	0.434236***	0.191784***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
QALY_sq	-0.004325***	-0.697661***	-0.038518***	-0.013868***	-0.002361***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
BOIQL	-0.027300***	0.000408	-0.070746***	-0.011632**	-0.019680***
	(0.000)	(0.992)	(0.000)	(0.024)	(0.000)
BOISU	0.009186***	0.149561***	-0.002667	0.032543***	-0.000020
	(0.000)	(0.000)	(0.825)	(0.000)	(0.992)
Log likelihood	-21489	-5148	-4013	-5138	-5346
Rho-squared	0.155	0.273	0.238	0.166	0.227
AIC	42987	10303	8034	10284	10700
BIC	43024	10335	8064	10315	10731
Mean VIF	4.85	6.95	5.26	6.18	8.04
stdQALY	16.026469***	18.131878***	15.670843***	17.369445***	15.342754***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdQALY_sq	-15.843892***	-17.441529***	-15.407104***	-22.189126***	-15.107985***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdBOIQL	-0.879914***	0.002041	-1.414921***	-0.465295**	-1.574378***
	(0.000)	(0.992)	(0.000)	(0.024)	(0.000)
stdBOISU	0.651972***	0.747806***	-0.053345	1.301718***	-0.001593
	(0.000)	(0.000)	(0.825)	(0.000)	(0.992)
Log likelihood	-19753	-5148	-4013	-5138	-5346
Rho-squared	0.223	0.273	0.238	0.166	0.227
AIC	39514	10303	8034	10284	10700
BIC	39550	10335	8064	10315	10731
Mean VIF	6.25	6.95	5.26	6.18	8.04
Observations	73,380	20,440	15,200	17,780	19,960

pval in parentheses; *** p<0.01, ** p<0.05, * p<0.1; QALY – QALY gains from new treatment; QALY_sq – QALY squared; BOIQL – QALY loss due to poor HRQoL; BOISU – QALY loss due to shorter life expectancy

3.4.3 QALY gains and life expectancy before treatment or end of life

As before, the QALY coefficient is positive and the QALY squared coefficient is negative and both effects are similar to previous models (Table 9). Survival before treatment (E) is negative and statistically significant across the models and improves their overall performance. The coefficients

on survival before treatment decline with increasing normal life expectancy. In regressions involving EOL with survival gains of 3 or more months rather than E, EOL is positive indicating respondents gave greater weight to shorter life expectancy before treatment when survival gains were greater than 3 months. However, as would be expected due to the loss of information when E is converted to a dummy variable, these models perform worse than equivalent models including E. The coefficients on EOL also decline with normal life expectancy except for the 40 year variant. Standardisation stabilises the coefficients across the questionnaire variants apart from the 40 year variant. The marginal rate of substitution of EOL is -1.84 QALYs using the 80 year variant and regression using non-standardised variables.

Table 9: QALY gains and life expectancy before treatment/ end of life

Variables	All	5yrs	20yrs	40yrs	80yrs
QALY	0.307033***	3.626240***	0.787416***	0.437654***	0.189601***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
QALY_sq	-0.004302***	-0.697631***	-0.038544***	-0.014124***	-0.002337***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
E	-0.024580***	-0.149347***	-0.043401***	-0.039460***	-0.010979***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	-21545	-5148	-4024	-5140	-5371
Rho-squared	0.153	0.273	0.236	0.166	0.224
AIC	43096	10301	8054	10287	10747
BIC	43123	10325	8077	10310	10771
Mean VIF	5.99	6.93	4.81	5.37	7.99
stdQALY	16.091462***	18.13120***	15.74831***	17.50618***	15.16804***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdQALY_sq	-15.94931***	-17.44077***	-15.41745***	-22.5982***	-14.9555***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdE	-1.162729***	-0.746736***	-0.868016***	-1.578411***	-0.878320***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	-19784	-5148	-4024	-5140	-5371
Rho-squared	0.222	0.273	0.236	0.166	0.224
AIC	39575	10301	8054	10287	10747
BIC	39602	10325	8077	10310	10771
Mean VIF	5.83	6.93	4.81	5.37	7.99
QALY	0.281260***	3.229699***	0.761621***	0.400385***	0.175159***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
QALY_sq	-0.003906***	-0.602180***	-0.036789***	-0.013910***	-0.002117***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EOL	0.609437***	0.606509***	0.375107***	0.576401***	0.314355***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	-21411	-5103	-4008	-5203	-5395
Rho-squared	0.158	0.280	0.239	0.156	0.220
AIC	42829	10213	8022	10411	10797
BIC	42857	10236	8045	10435	10820
Mean VIF	5.24	7.78	4.72	5.26	7.44
stdQALY	14.748108***	16.148494***	15.232415***	16.015386***	14.012740***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdQALY_sq	-14.409639***	-15.054491***	-14.715728***	-22.255325***	-13.549861***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EOL	0.487499***	0.606509***	0.375107***	0.576401***	0.314355***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	-19806	-5103	-4008	-5203	-5395
Rho-squared	0.221	0.280	0.239	0.156	0.220
AIC	39619	10213	8022	10411	10797
BIC	39646	10236	8045	10435	10820
Mean VIF	5.54	7.78	4.72	5.26	7.44
Observations	73,380	20,440	15,200	17,780	19,960

pval in parentheses; *** p<0.01, ** p<0.05, * p<0.1; QALY – QALY gains from new treatment; QALY_sq – QALY squared
E –life expectancy before treatment; EOL – life expectancy before treatment ≤2 years and with survival gains≥ 3 months.

3.4.4 QALY gains and HRQoL and life expectancy before treatment

Life expectancy before treatment (E) is negative, statistically significant and a similar magnitude to those reported in models with only E (Table 10). Standardised HRQoL before treatment (stdH) is positive and statistically significant before treatment. E and H both appear in the QALY so the coefficients on their own cannot be assessed. Further assessment indicates that the effect of E is

negative within the range of likely levels of E (i.e. less than 100 years) while H is positive. This indicates that respondents were more likely to choose profiles with lower life expectancy but with higher HRQoL which supports the split BOI finding for burden from HRQoL but is in contrast to the results where there are no QALY terms (Table 6). Standardisation minimises differences across the questionnaire variants apart from the 40 year variant.

Table 10: QALY gains with life expectancy and HRQoL before treatment

Variables	All	5yrs	20yrs	40yrs	80yrs
QALY	0.295534***	3.590566***	0.765465***	0.431436***	0.184972***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
QALY_sq	-0.004176***	-0.690254***	-0.038048***	-0.013868***	-0.002303***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
E	-0.023122***	-0.146742***	-0.041931***	-0.038626***	-0.010430***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdH	0.792473***	0.128931	0.775565***	0.251165***	0.639825***
	(0.000)	(0.135)	(0.000)	(0.003)	(0.000)
Log likelihood	-21348	-5147	-3987	-5136	-5341
Rho-squared	0.161	0.273	0.243	0.167	0.228
AIC	42705	10301	7982	10280	10689
BIC	42742	10322	8012	10311	10721
Mean VIF	4.75	5.47	3.86	4.34	6.39
stdQALY	15.70341***	17.95283***	15.3093***	17.25744***	14.79776***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdQALY_sq	-15.61605***	-17.25636***	-15.21917***	-22.18838***	-14.74235***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdE	-1.112958***	-0.733708***	-0.838610***	-1.545033***	-0.834435***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdH	0.478945***	0.128931	0.775565***	0.251165***	0.639825***
	(0.000)	(0.135)	(0.000)	(0.003)	(0.000)
Log likelihood	-19721	-5147	-3987	-5136	-5341
Rho-squared	0.225	0.273	0.243	0.167	0.228
AIC	39450	10301	7982	10280	10689
BIC	39487	10322	8012	10311	10721
Mean VIF	4.62	5.47	3.86	4.34	6.39
Observations	73,380	20,440	15,200	17,780	19,960

pval in parentheses; *** p<0.01, ** p<0.05, * p<0.1; QALY – QALY gains from new treatment; QALY_sq – QALY squared; H and E – HRQoL and life expectancy before treatment

Notes: stdH i.e. H/100 is used in the unstandardized QALY model to reflect the fact that H/100 is used to build the QALY. This has no effect on the direction of the coefficient; it just increases the magnitude by a factor of 100.

3.4.5 Comparison of model performance

The best fitting models based on the log-likelihood, rho-squared, BIC and AIC are the ones with simple terms for HRQoL and life expectancy before and after treatment models (H, H+Q, E, E+S). When comparing the QALY models, which are more relevant for policy, the best fitting models are those with HRQoL and life expectancy before treatment rather than BOI. Splitting BOI resulted in better performing models using all model performance criteria, yet the interpretation and suitability

of the split variables is questionable. Standardisation improves the performance of models when results are pooled across the questionnaire variants, but makes no difference within variant as would be expected. Mean variance inflation factors (VIF) across all the models were less than 10 indicating that there was no collinearity. Individual VIFs across the terms (not reported) show that the only terms with VIF factors higher than 5 were squared terms i.e. QALY squared which is expected. All other terms such as BOI and E had VIFs of approximately 1.

3.4.6 Robustness of results

There were concerns that respondents who did not understand or engage with the survey may have had an impact on results. First, the consequences were examined of excluding the following: 279 individuals who reported they had difficulty understanding the DCE questions; 208 individuals who took less than 5 minutes or more than 60 minutes to complete the survey; and 23 individuals who chose the same option for all their DCE questions. These exclusions did not have consequences for the significance and direction of the QALY, BOI and EOL coefficients, though there were variations in magnitude for all of them. The impact of excluding the first question for each individual to test for learning effects was also examined, and no impact on the coefficients in terms of significance and direction was found but there were small changes in magnitude.

Second, the consequences of excluding individuals from the analyses who arguably misunderstood the DCE task were examined. 1,442 respondents were excluded who chose a profile that had a larger number of total lifetime QALYs after treatment, but smaller QALY gain from treatment and lower BOI before treatment than the other profile. These respondents were excluded on the grounds that they misunderstood the task as they chose the profile they thought was best or that they wanted to live in, not the profile with the patients who were most deserving of treatment in terms of size of QALY gain and BOI. Some respondents did not answer a question that could be used to implement this exclusion criteria, and these 369 respondents remained in the analysis, though it is possible that some of these respondents may have also misunderstood the DCE task. Appendix 2, Table 1 provides summary characteristics of the sample of excluded respondents and the remaining sample of 2,247 respondents. In comparison to the remaining sample, the sample of respondents who misunderstood the DCE task had a higher proportion of older respondents, respondents who were retired or long-term sick and whose highest level of education was secondary school, respondents in poorer health and respondents who found the DCE and attitudinal questions difficult. Appendix 2, Table 2 provides responses to the practice questions, demonstrating that in practice question 1, a larger proportion of respondents (91.4% in full sample compared to 94.0% for the sample excluding

respondents who misunderstood the DCE task) chose to treat the group with the highest treatment gain. The results for practice question 2 are more striking, where the proportion of all respondents choosing to treat the patients with higher BOI was 50.8%, yet following the exclusion of respondents who misunderstood the DCE task the proportion choosing to treat the patients with higher BOI was 63.5%.

All regressions reported in section 3.4 were re-estimated excluding respondents who misunderstood the DCE task and are reported in Table 11 and Appendix 2. The results are consistent with the main results reported above for significance and direction of the QALY and EOL coefficients, though there were variations in magnitude for all of them. However, for BOI the coefficients are positive and significant at the 1% level and substantially larger in magnitude (Table 11). When BOI is split between premature mortality and morbidity the coefficients are always positive and are significant with the exception of the 80 year variant for BOI from HRQoL (Table 11). In addition the regressions estimated with QALY, QALY squared, life expectancy before treatment and standardised HRQoL before treatment have the expected significance and direction of coefficients where respondents are more likely to choose profiles with lower life expectancy and HRQoL before treatment. Overall these results suggest that there is a preference for BOI amongst respondents who seemed to have understood the DCE task.

Table 11: QALY and BOI, and QALY and split BOI, excluding respondents who misunderstood the DCE task

Variables	All	5 years	20 years	40 years	80 years
QALY	0.327731***	4.296049***	0.977076***	0.534826***	0.201342***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
QALY_sq	-0.004628***	-0.871767***	-0.050522***	-0.018881***	-0.002557***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
BOI	0.071495***	0.686080***	0.284999***	0.130007***	0.035929***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	-12923	-3239	-2086	-2698	-3330
Rho-squared	0.170	0.326	0.282	0.217	0.247
AIC	25851	6485	4178	5402	6666
BIC	25878	6507	4199	5424	6688
Mean VIF	5.61	6.99	5.04	5.23	7.42
stdQALY	18.565652***	21.480247***	19.541523***	21.393033***	16.107366***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdQALY_sq	-19.338888***	-21.794173***	-20.208719***	-30.210193***	-16.363127***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdBOI	3.946421***	3.430399***	5.699972***	5.200264***	2.874328***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	-11470	-3239	-2086	-2698	-3330
Rho-squared	0.264	0.326	0.282	0.217	0.247
AIC	22945	6485	4178	5402	6666
BIC	22972	6507	4199	5424	6688
Mean VIF	5.79	6.99	5.04	5.23	7.42
QALY	0.380000***	4.396137***	1.063459***	0.601136***	0.232766***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
QALY_sq	-0.005393***	-0.891345***	-0.054045***	-0.019291***	-0.003020***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
BOIQL	0.019246***	0.506098***	0.143250***	0.056376***	0.005484
	(0.000)	(0.000)	(0.000)	(0.000)	(0.126)
BOISU	0.063979***	0.735917***	0.275993***	0.123955***	0.030458***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	-12656	-3220	-2027	-2590	-3250
Rho-squared	0.187	0.330	0.302	0.248	0.265
AIC	25319	6448	4062	5188	6509
BIC	25354	6478	4091	5216	6539
Mean VIF	5.08	7.05	5.49	6.07	8.30
stdQALY	20.632141***	21.980684***	21.269172***	24.045446***	18.621290***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdQALY_sq	-21.676245***	-22.283617***	-21.618083***	-30.866060***	-19.325591***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdBOIQL	1.718467***	2.530489***	2.865009***	2.255054***	0.438691
	(0.000)	(0.000)	(0.000)	(0.000)	(0.126)
stdBOISU	3.847949***	3.679584***	5.519870***	4.958199***	2.436641***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	-11190	-3220	-2027	-2590	-3250
Rho-squared	0.282	0.330	0.302	0.248	0.265
AIC	22389	6448	4062	5188	6509
BIC	22424	6478	4091	5216	6539
Mean VIF	6.42	7.05	5.49	6.07	8.30
Observations	44,940	13,860	8,380	9,940	12,760

pval in parentheses; *** p<0.01, ** p<0.05, * p<0.1; QALY – QALY gains from new treatment; QALY_sq – QALY squared; BOI – QALY loss from condition with current treatment; BOIQL – QALY loss due to poor HRQoL; BOISU – QALY loss due to shorter life expectancy

3.5 Attitudinal questions

Results from the attitudinal questions are reported in Table 12.

Table 12: Responses to the attitudinal questions

	Normal life expectancy	5	20	40	80	All
Question	**Response N**	**1022**	**760**	**889**	**998**	**3669**
	BOI					
1	The NHS should give priority to treating patients who are very ill	40.3%	41.7%	40.9%	40.2%	40.7%
	The NHS should give the same priority to treating all patients who are ill, regardless of how ill they are	59.7%	58.3%	59.1%	59.8%	59.3%
2	The NHS should give priority to treating patients who are very ill and will die early because of their illness	42.1%	41.6%	43.3%	42.8%	42.5%
	The NHS should give the same priority to treating all patients who are ill, regardless of how ill they are or when they will die	57.9%	58.4%	56.7%	57.2%	57.5%
3	The NHS should always give priority to treating patients who are very ill and will die early because of their illness, even if they only get a small amount of benefit from treatment	8.1%	9.9%	10.8%	9.6%	9.5%
	The NHS should give priority to treating patients who are very ill and will die early because of their illness, but only if they get a large amount of benefit from treatment	46.7%	44.5%	41.4%	44.7%	44.4%
	The NHS should give the same priority to treating all patients, regardless of how ill they are or when they will die	45.2%	45.7%	47.8%	45.7%	46.1%
8	The NHS should give priority to treating patients who are very ill and will die early because of their illness	9.3%	10.3%	12.8%	10.2%	10.6%
	The NHS should give priority to treating patients who will get the largest amount of benefit from treatment	47.2%	45.5%	43.0%	42.8%	44.6%
	The NHS should give the same priority to treating all patients	43.5%	44.2%	44.2%	47.0%	44.8%
	EOL					
4	The NHS should give priority to extending the life of patients who are expected to die soon, even if this is the natural end of their life	5.7%	6.8%	7.3%	5.5%	6.3%
	The NHS should give priority to patients expected to die soon, but only if it means they die before the natural end of their life	38.6%	38.0%	37.7%	39.1%	38.4%
	The NHS should give the same priority to treating all patients, regardless of how ill they are or when they will die	55.7%	55.1%	55.0%	55.4%	55.3%
5	The NHS should give priority to extending the life of patients who are expected to die soon, even if this means they live in very poor health	3.1%	4.3%	4.2%	4.0%	3.9%
	The NHS should give priority to extending the life of patients who are expected to die soon, but only if they would live in a reasonable level of health	56.2%	57.0%	56.9%	55.7%	56.4%
	The NHS should give the same priority to treating all patients, regardless of how ill they are or when they will die	40.7%	38.7%	38.9%	40.3%	39.7%
7	The NHS should give priority to extending the life of patients expected to die soon	11.8%	12.0%	13.8%	10.7%	12.0%
	The NHS should give priority to treating patients who will get the largest amount of benefit from treatment	88.2%	88.0%	86.2%	89.3%	88.0%
	Therapeutic Improvement					
6	The NHS should give priority to treatments that give a large amount of benefit to a small number of patients	8.8%	8.2%	8.9%	6.8%	8.1%
	The NHS should give priority to treatments that give a small	8.3%	11.3%	10.6%	8.3%	9.5%

	Normal life expectancy	5	20	40	80	All
	amount of benefit to a large number of patients					
	The NHS should consider the amount of benefit a treatment gives overall, rather than considering how it is shared out among different numbers of patients	82.9%	80.5%	80.5%	84.9%	82.4%
	Combined					
9	The NHS should give priority to treating patients who are very ill and will die early because of their illness	12.8%	12.6%	14.4%	12.4%	13.1%
	The NHS should give priority to treating patients who will get the largest amount of benefit from treatment	52.2%	53.8%	52.8%	50.3%	52.1%
	The NHS should give priority to treating patients who will live for a long time and be in good health after treatment	35.0%	33.6%	32.8%	37.3%	34.8%

3.5.1 Burden of illness (BOI)

Just under half answered that the NHS should give priority to treating patients who are very ill (i.e. higher BOI) or those who are very ill and will die early (40.7% and 42.5% respectively) (Table 12, Q1 and 2), rather than giving the same priority to all patients. Just 9.5% of respondents indicated that the NHS should give priority to treating patients who are very ill and will die early (i.e. higher BOI) if they only get a small amount of benefit from treatment compared to 44.4% of respondents who thought that the NHS should give priority to treating patients with a higher BOI but only if they get a large benefit from treatment, and 46.1% of respondents thought that the NHS should give the same priority to all patients (Q3). This implies that the value placed on BOI depends on the gain. 10.6% of respondents thought the NHS should give priority to patients with high BOI (i.e. very ill and will die early) compared to 44.6% of respondents who said that the NHS should give priority to patients with the highest treatment gain and 44.8% who thought the same priority should be given to all patients (Q8). Overall, this indicates that treatment gain was preferred over BOI but marginally more respondents preferred to give equal priority to all patients.

Question 2 was used to assess whether the attitudes were consistent with the DCE results. The regression models were re-estimated using variables of QALY, QALY squared and BOI for subsamples generated using their responses to question 2. As would be expected, the regression estimated for the subsample who showed a preference for BOI, also had positive coefficients for BOI (Appendix 3 Table 1) while those who did not show a preference for BOI had negative coefficients.

3.5.2 End of life (EOL)

Very few respondents indicated that the NHS should give priority to patients expected to die soon if this is the natural end of their life (6.3%), but 38.4% of respondents thought that the NHS should give priority to patients expected to die soon if they die before the natural end of their life (Q4).

However, more than half (55.3%) thought the NHS should give the same priority to all patients. So EOL matters to nearly half of the respondents and then only if it is before normal life expectancy. In question 5, only 3.9% of respondents thought the NHS should give priority to extending the life of patients expected to die soon even if they live in poor health. There were 56.4% of respondents who indicated that the NHS should give priority to extending the life of patients expected to die soon but only if it means they live in a reasonable level of health, while 39.7% of respondents thought that the NHS should give the same priority to all patients. In this latter case, EOL was important but only where the patients lived in reasonable health. When the choice was between EOL and treatment gain, 88% of respondents chose to give priority to largest treatment gain over EOL (Q7).

Question 5 was used to assess whether the attitudes were consistent with the DCE results, where the regression involving QALY, QALY squared and EOL was re-estimated on subsamples generated using their responses to question 5. As would be expected, the EOL coefficient was larger for those who show a preference towards EOL compared to those who did not (Appendix 3 Table 2).

3.5.3 Therapeutic improvement (TI)

Only 8.1% of respondents thought that the NHS should give priority to treatments with large treatment gain for a small number of patients. There were 82.4% of respondents who indicated that the NHS should consider the amount of benefit overall rather than considering how it is shared out among different numbers of patients, again suggesting a preference for QALY maximisation.

Question 6 was used to assess whether the attitudes were consistent with the DCE results. Regressions involving QALY, QALY squared and BOI were re-estimated on subsamples generated using question 6. There were small differences across the 3 groups representing those who had a preference for concentration, dispersion or no preference for either (Appendix 3, Table 3). However, given the small numbers of respondents in the first two groups, these results may be driven by the size and composition of profiles in the smaller samples so they should be treated with caution.

Overall the responses to these questions indicated that most respondents believed that the NHS should give preference to the group with the largest treatment gain over BOI or EOL. The results suggest some support for BOI, though this depended on the size of treatment gain and some support for EOL provided they live at a reasonable level of health. There is very little support for TI. However it must be remembered that these are dichotomous questions and do not permit trading-off between attributes. A large proportion of respondents consistently indicated that the same priority should be given to all patients indicating that they were not QALY maximisers.

3.5.4 Robustness of results

Comparing responses to the attitudinal questions for respondents who misunderstood the DCE analysis (see section 3.4.6) and the remaining respondents showed that a large proportion of respondents who misunderstood the DCE task chose that the NHS should give the same priority to treating all patients, regardless of how ill they are or when they will die. In contrast, the percentage of respondents expressing a preference for BOI and EOL was higher for the remaining sample of respondents than for respondents who misunderstood the DCE task. The responses to the attitudinal questions were therefore consistent with the regression results excluding the respondents who seemed to have misunderstood the DCE task.

4 DISCUSSION

4.1 Summary

This was the first study to examine societal preferences over BOI alongside TI and EOL. It was a large DCE survey using an existing online panel drawn from the general population. The results confirmed that respondents took into account the profile information provided in the text in the expected way i.e. they were more likely to choose profiles with lower life expectancy and HRQoL before treatment (severe health) and profiles with higher life expectancy and HRQoL after treatment (larger gains). The key findings in policy terms were that survey participants preferred to treat patients who had larger QALY gains but at a diminishing rate, meaning there was no support for TI. They also preferred to treat patients who had a shorter life expectancy or who were at the end of life (EOL). The results for BOI were less robust across versions of the questionnaire but suggested some modest support for BOI. Regressions excluding respondents who seemed to have misunderstood the DCE task found larger and more robust support for BOI. The attitudinal questions seemed to support the regression results for QALY gains and BOI although less so for EOL. The attitude questions also indicated that many respondents preferred to treat all patients the same, though the meaning of this response is unclear.

4.2 QALY gains and therapeutic improvement

The results of this survey indicated that respondents tend to choose to treat the group with the larger treatment gain, but they did not support the notion of TI (QALY squared) set out in the value based pricing consultation document.[DH, 2010] The negative QALY squared term indicated a diminishing marginal gain with each increase in QALYs. All other things being equal, there was a

switching point where the impact of further QALY gains became negative overall. This indicated a point where there was a preference for smaller gains over larger gains. The precise point of this switch was specific to each questionnaire variant. Although the model with the QALY squared term fitted the data better statistically, it may not be representative of societal preferences but rather may be as a result of the study design. QALY gains were heavily skewed towards smaller gains with few scenarios with large QALY gains and this may have had an effect on the results. However, though it is not directly comparable in terms of the attributes included, Lancsar et al.[2011] found QALY gains to have a positive and statistical impact but again at a declining rate, as did a recent study by Norman et al.,[2012] with both rejecting the notion of TI. The application of this declining weight on QALY gain could have a major impact on health care policy making.

4.3 End of life

The regression results showed that there was support for EOL across the regression models, with evidence for a preference to treat those who were at the end of their life. Although responses to the attitude questions cast some doubt on this finding, as this was not a view held by the majority, testing the DCE results based on attitudes indicated that those whose attitudes supported EOL also had larger coefficients for EOL in the DCE. While size of treatment effect was more important than EOL, it does not mean that EOL had no value. It should be noted that the current NICE criteria for EOL also states that the condition should be rare,[NICE, 2009] but rarity was not examined in this study.

4.4 Burden of illness

The findings from the attitude questions and the DCE suggested some support for BOI, though the findings were not consistent across all models estimated using all responses. The responses to practice question 2 found little support for BOI across all respondents with between 46.8% and 54.3% of respondents across variants choosing the option with higher BOI. However, excluding respondents who misunderstood the DCE task indicated that remaining respondents had a clear preference for BOI (63.5%). This provided support that BOI did have a positive impact overall.

BOI provided a broader notion of severity than previous research since it incorporated the impact on QoL and length of life over the patient's future life. The impact of BOI was quite modest, though significant, in three of the four normal life expectancy groups when estimated using all responses,

but was always larger and significant in regressions across all variants excluding respondents who misunderstood the DCE task.

Splitting the BOI term indicated that the small positive coefficient in the BOI term was a result of different effects of burden from poor HRQoL (BOIQL) and burden from shorter life expectancy (BOISU). The latter was as expected and was positive indicating that respondents were more likely to choose profiles with higher QALY loss from shorter life expectancy. The negative coefficients on BOIQL would seem to contradict the severity argument (which implies that lower pre-treatment HRQoL should increase the weight given to a given BOIQL), since the negative coefficients implied respondents preferred to treat those in less severe health in HRQoL terms. However, BOIQL was not solely attributable to HRQoL, since it was the result of the product of pre-treatment HRQoL and survival. BOIQL can increase due to either a reduction in HRQoL (i.e. H) or an increase in pre-treatment survival. Therefore, BOIQL cannot be seen as a test of the conventional severity argument. Equally this suggests that BOI cannot be split into a HRQoL effect and a survival effect, since by definition BOI is composed of both. Regressions estimated excluding respondents who misunderstood the DCE task had positive and significant coefficients for BOISU and BOIQL (except the coefficient for BOIQL was not significant for the 80 year variant).

The coefficients for BOI were not consistent across the four normal life expectancies and did not seem to exhibit an obvious pattern. Conversely, the coefficient for the QALY term declined with normal life expectancy. Standardising the BOI and QALY terms did improve the performance of the model in the pooled data, but it did not stabilise the BOI coefficients across the life expectancies. The main results do not support the notion of a proportional shortfall approach to BOI proposed by Stolk et al.[2005] However, the regressions estimated excluding respondents who misunderstood the DCE task are more consistent with the proportional shortfall approach as the coefficients for the standardised BOI terms are more similar across the variants. The experiment by Stolk and colleagues was different to the DCE used here since it assumed a magic pill that restored someone to full health and normal life expectancy.

4.4.1 Life expectancy, end of life and health related quality of life

The results from the models involving HRQoL before (H) and after (HQ) treatment and life expectancy before (E) and after (ES) treatment provided the best performing models. In contrast to above, respondents choose profiles with lower HRQoL before treatment. Further analysis to assess this based on life expectancy and HRQoL before treatment (E and H) indicated that when included

with the QALY term, E remained negative but H was positive. Estimating the overall impact of E and H in that model (since they also appear in the QALY terms), resulted in them remaining negative and positive respectively. This may mean that respondents did not take into account severity in HRQoL in the expected way. However, there were no scenarios where conditions A and B differed solely on the basis of HRQoL before treatment, so it is difficult to make any conclusions regarding the impact of this on preferences. This would seem to contradict the evidence on the impact of severity defined in terms of low pre-treatment HRQoL on social preferences, but many of these used different methods and were often substantially smaller in scale than this one.[Shah, 2009] Other recent large scale studies with members of the general population, including Lancsar et al.[2011] in the UK and Norman et al.[2012] in Australia found respondents were less likely to choose patient groups with a higher HRQoL loss which is consistent with the current findings. However, when respondents who misunderstood the DCE task were excluded from the analyses, coefficients for H and E were both negative and significant, suggesting that the results in the main analysis may have been due to the inclusion of respondents who misunderstood the DCE task.

4.5 Normal life expectancy (N)

Overall, the results indicated that normal life expectancy in the profiles (N) had an effect on the results with decreasing coefficient size on gains in life years as normal life expectancy increased. This effect was also present on life expectancy without treatment and QALY gains and BOI to a lesser extent. The largest impact of varying life expectancy was between 5 years and 20 years. Standardisation indicated that scale may have been an issue for the QALY and survival terms, and BOI in regressions excluding respondents who seemed to have misunderstood the DCE task, since much of the difference was removed. However, the standardisation of the variables did not completely eradicate this effect, meaning that other factors may have affected the results and absolute values may have continued to have a role. This is an important finding as it indicates that framing may affect results even when respondents see a single life expectancy as has been used in other studies that have followed a similar approach (e.g. Shah et al. 2012).

4.6 Existing literature

Since the start of this project two large studies have been conducted examining preferences for value based pricing. These recently conducted surveys in the literature support the current findings for BOI and TI but not for EOL. Linley and Hughes[2013] conducted an online survey of 4,118 UK general population asking respondents how they would prefer NHS money to be spent across two

groups of patients, where a scale was provided that meant if they chose to treat more patients in one group fewer patients were treated in the other group. Attributes were varied across the two groups of patients including severity of disease and EOL. The sample was split into two cohorts where further questions examining preferences were used. The first cohort was asked the same question but where treatment improves health a little for patients with severe disease but considerably for patients with moderate disease. The second cohort was asked the same question but treatment costs differ across the groups where treating patients with the severe disease has double costs in comparison to patients with moderate disease. The study found that there was a societal preference for the allocation of NHS funds to treat patients with severe rather than moderate disease, but not for treating patients with EOL. This is contrary to the results found in the current survey regarding EOL. The framing of the Linley and Hughes[2013] survey was very different to the framing of the current survey, and it is possible that this influenced the results. In particular the first cohort were asked a question combining size of QALY gain and severity or EOL, and the second cohort were asked to consider costs, whereas costs are not mentioned in the current survey.

In a small exploratory study (n=50) to assess the impact of EOL, Shah et al.[2013] found weak support for treating EOL patients while a larger follow-on study indicated little support for EOL.[Shah et al, 2012] The difference in preferences for EOL for the current survey and the larger follow-on survey is surprising given that both used a similar question format to elicit preferences (the Shah et al. 2012 survey used study 2 described above to inform the design of their survey), though framing effects obviously remain. The Shah et al.[2012] study involved an online survey of 3,969 UK respondents asking which of two hypothetical patients should be treated, where the health service has enough funds to treat one patient but not both. Respondents were provided information on the life expectancy and quality of life of the patients without treatment, and life expectancy and quality of life gains from treatment. In contrast to the current survey, respondents were not provided with any information about normal life expectancy and quality of life if the patient did not have the condition. Whilst the study indicated little support for EOL, the size of the QALY gains were found to be important and this is consistent with the findings regarding QALY gains in the current study.

4.7 Limitations

A concern with this survey is the use of an online sample, since it may exclude groups in society such as the computer illiterate or those unable to access a computer. The use of an online panel means that respondents have stated that they are willing to regularly answer online surveys, and this also makes them unrepresentative of computer users. As the respondents receive points for every survey

they complete that can be exchanged for goods this also may lead to the reasoning for answering the survey to be questioned, and may again indicate that these respondents are not representative of the UK. The importance of these selection processes for the responses obtained in the survey is not known. However, the sample was recruited using a nationally representative quota for age and gender and for these characteristics the sample is nationally representative.

Preparatory studies undertaken before the main survey suggested that some respondents failed to understand the tasks set. The main survey was therefore designed to minimise this problem and comprised: an introduction video, 2 practice questions with feedback and profiles that included pictures to aid understanding. Respondents' views of the survey indicated that the majority of respondents did not find either the DCE tasks or the attitudinal questions difficult. However there was evidence to suggest that respondents did not always understand the DCE tasks, where respondents failed to select to treat the patients who were most deserving of treatment. Some respondents failed to understand that the task asked them to choose which patient group to treat, as they chose the better health profile in overall QALY terms, not the profile with the patients who were most deserving of treatment in terms of BOI and QALY gains.

Respondents who chose a profile that had a higher number of total lifetime QALYs after treatment, but smaller QALY gain from treatment and lower BOI before treatment than the other profile were therefore excluded in robustness analyses. The exclusion of these respondents could be criticised on the grounds that it was imposed after the data were collected. Although a large number of respondents are excluded using this criterion, this is not a reason for including respondents who misunderstood the tasks as their responses are not random and show a preference for the better profile rather than the patients they wish to treat. However, it is possible that some of these respondents did understand the task and were expressing a genuine preference, though this is unlikely given the response. Qualitative work would be required to ascertain the full nature and extent of the understanding of the task and this was not undertaken in this study. The exclusion of individuals who misunderstood the DCE task in robustness analyses did not affect the significance and direction of terms representing QALY gains, TI and EOL, but for BOI the coefficients were larger and the sign consistent across the variants.

While there is no explicit reference to age in the questions, it could be argued that respondents would be able to estimate the age of the recipients of the treatment from the normal life expectancies. However, this did not come up in preparatory study 3, a small qualitative survey,

where respondents were asked what they were thinking about as they were completing the task. Furthermore, each respondent only faced one normal life expectancy and so the survey did not draw attention to this feature. The results by subgroup did find differences across the four life expectancies, which were substantially reduced when standardisation was carried out. This indicates that life expectancy influenced results but it is not possible to verify whether respondents were thinking about age or other factors such as end of life when they were faced with the fixed life expectancy.

The study assumes that respondents have a zero time preference, but it is possible that this assumption may not be correct. Whilst societal time preferences may not be as large as individual time preferences, this remains a limitation of the study.

The attitudinal questions suffer from the limitation that they are dichotomous and do not address trade-offs between attributes. The format of these questions may mean that respondents state that all patients should be treated the same while their response may have been different if a trade-off had been required. In contrast the DCE tasks require a choice and do not allow respondents to state that they are indifferent, or in effect that they wish to treat all respondents the same. However, it is expected that these respondents will randomly choose to treat either of the patient groups and hence this will not affect the regression results.

5 CONCLUSION

This study provides the first attempt to operationalise the concept of BOI which combines the conventional notion of severity (in terms of low HRQoL) with survival (using QALY loss from the condition) to provide an indicator of the severity of a condition. It also provides evidence on societal preferences for EOL and the size of the gain (TI). The results indicate some support for BOI as a consideration when weighting QALYs, but it seems to be driven by burden from premature mortality rather than morbidity. The evidence did not support the idea of TI. There was robust and consistent support for EOL in general. Overall there seems to be a strong preference for larger QALY gain but at a diminishing rate. These results support the argument that the social value of a QALY is not equal between recipients and depends on both the burden of their condition and expected survival.

6 REFERENCES

Akaike H. (1973) Information Theory and an Extension of the Maximum Likelihood Principle. In: B N Petrov and F Csaki, eds. Second International Symposium on Information Theory. Budapest: Akademiai Kiado, 267–281.

All Wales Medicines Strategy Group (AWMSG). (2012) AWMSG summary guidelines for appraising medicines. Llandough: AWSMG.

Ara R, Brazier J. (2011) Using health state utility values from the general population to approximate baselines in decision analytic models when condition specific data are not available. Value in Health 14: 539-545.

Baker R, Bateman I, Donaldson C, Jones-Lee M, Lancsar E, Loomes G, Mason H, Odejar M, Pinto Prades JL, Robinson A, Ryan M, Shackley P, Smith R, Sugden R, Wildman J. (2010) Weighting and valuing quality-adjusted life-years using stated preference methods: preliminary results from the Social Value of a QALY project. Health Technology Assessment, 14: 27.

Brazier J, Ratcliffe J, Salomon J, Tsuchiya A. (2007) The measurement and valuation of health benefits for economic evaluation. Oxford University Press, Oxford.

Carlsson F, Martinsson P. (2003) Design techniques for stated preference methods in health economics. Health Economics 12: 281-294.

Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, Devlin N, Smith PC, Sculpher M. (2013) Methods for the estimation of the NICE cost effectiveness threshold. Centre for Health Economics Research Paper 81, University of York (http://www.york.ac.uk/media/che/documents/papers/researchpapers/CHERP81_Methods_estimation_NICE_costeffectiveness_threshold.pdf).

Department of Health. (2010) Value based pricing: impact assessment. London: Department of Health (available at: http://www.dh.gov.uk/en/Consultations/Liveconsultations/DH_122760).

Department of Health. (2012) Implementing a ban on age discrimination in the NHS – making effective, appropriate decisions. London: Department of Health (available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/212944/ban-on-age-discrimination.pdf).

Devlin N, Parkin D**.** (2004) Does NICE have a cost effectiveness threshold and what other factors influence its decisions? A binary choice analysis. Health Economics 13: 437-452.

Dolan P, Edlin R, Tsuchiya A. (2008) The relative societal value of health gains to different beneficiaries. Final report - NICE Social QALY project.

Green C. (2009) Investigating public preferences on 'severity of health' as a relevant condition for setting health care priorities. Social Science and Medicine 68: 2247-2255.

Hannson LF, Norheim OF, Ruyter KW. (1994) Equality, explicitness, severity and rigidity: The Oregon Plan evaluated from a Scandinavian perspective. The Journal of Medicine and Philosophy 19: 343-366.

Johannesson M. (2001) Should we aggregate relative or absoute changes in QALYs? Health Economics; 10, 573-577.

Kind P, Hardman G, Macran S. (1999) UK population norms for EQ-5D. Centre for Health Economics Discussion Paper Series, University of York.

Keetharuth A, Brazier J, Tsuchiya A. (Forthcoming) Review on the Social Value of a QALY. EEPRU Research Report.

Kuhfeld WF. (2005) Marketing Research Methods in SAS. SAS Institute Inc.: Cary.

Lancsar E, Wildman J, Donaldson C, Ryan M, Baker R. (2011) Deriving distributional weights for QALYs through discrete choice experiments. Journal of Health Economics, 30, 466-478.

Luce RD. (1959) Individual choice behaviour: a theoretical analysis. John Wiley & Sons, Inc. New York.

McFadden D. (1974) Conditional logit analysis of qualitative choice behaviour. In Zarembka P, ed. Frontiers in econometrics. Academic Press, New York pp. 105-142.

Murray CJL, Lopez AD. (1996) The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Published by the Harvard School of Public Health on behalf of the World Health Organization and the World Bank.

National Institute for Health and Clinical Excellence (NICE). (2009) Appraising life-extending, end of life treatments. London, NICE.

National Institute of Health and Care Excellence (NICE). (2013) Guide to the methods of technology appraisal 2013. London: NICE.

Nord E. (1993) The trade-off between the severity of illness and treatment effects in cost-value analysis of health care. Health Policy 24: 227-238.

Nord E. (2005) Concerns for the worse off: fair innings versus severity. Social Science and Medicine 60: 257-263.

Norman R, Hall J, Street D, Viney R. (2012) Efficiency and equity: a stated preference approach. Health Economics (DOI.1002/hec).

Rowen D, Brazier J, Keetharuth A, Tsuchiya A, Hole AR, Shackley P, Robinson A. (Forthcoming *a*) Preparatory studies for eliciting societal preferences for Value-Based Pricing. EEPRU Research Report.

Rowen D, Brazier J, Keetharuth A, Tsuchiya A, Mukuria C. (Forthcoming *b*) Comparison of an online survey and face-to-face interviews for eliciting societal preferences for Value-Based Pricing. EEPRU Research Report.

Schwarz G. (1978) Estimating the dimension of a model. Annals of Statistics 6:461-464.

Salomon *et al*. (2012) “Common values in assessing health outcomes from disease and injury: disability weights measurement study for the GBD 2010” Lancet 380: 2129-43.

Scottish Medicines Consortium (SMC). (2013) Guidance to Manufacturers for Completion of New Product Assessment Form (NPAF). Glasgow: SMC.

Shah KK. (2009) Severity of illness and priority setting in healthcare: A review of the literature. *Health Policy*, 93: 77-84.

Shah KK, Tsuchiya A, Hole AR and Wailoo A. (2012) Valuing health at the end of life: A stated preference discrete choice experiment. Report by NICE DSU (http://www.nicedsu.org.uk/DSU%20End%20of%20Life%20full%20report%20-%20version%203%20_Dec%202012_.pdf).

Shah KK, Tsuchiya A, and Wailoo A. (2013) Valuing health at the end of life: an empirical study of public preferences. The European Journal of Health Economics 1-11.

Stolk EA, Pickee SJ, Ament AH, Busschbach JJ. (2005) Equity in health care prioritisation: an empirical inquiry into social value. Health Policy 74: 343-355.

Weinstein MC. (1988) A QALY is a QALY--or is it? Journal of Health Economics 7: 289-290.

Williams A. (1997) Intergenerational equity: an exploration of the 'fair innings' argument. Health Economics 6:117-32.

7 Appendix

Appendix 1: Standardised descriptive statistics

Appendix 1 Table 1: Mean standardised values across the questionnaire variants

version	stdH	stdQ	stdE	stdS	stdHQ	stdES	stdQALY	stdBOI	stdQLYSU	stdQLYQL	stdBOISU	stdBOIQL
5 yrs	0.41638	0.18271	0.24111	0.14816	0.59909	0.38926	0.1346	0.90135	0.088362	0.046236	0.75889	0.14245
20 yrs	0.41111	0.1719	0.18105	0.12305	0.583	0.3041	0.10196	0.92588	0.070834	0.031128	0.81895	0.10693
40 yrs	0.41476	0.17429	0.21664	0.06254	0.58905	0.27918	0.07529	0.91218	0.036562	0.038731	0.78336	0.12882
80 yrs	0.41159	0.17132	0.19499	0.12932	0.58291	0.32431	0.10637	0.91686	0.072397	0.033973	0.80501	0.11185

Appendix 1 Table 2: Range (maximum - minimum) standardised values across the questionnaire variants

version	stdH	stdQ	stdE	stdS	stdHQ	stdES	stdQALY	stdBOI	stdQLYSU	stdQLYQL	stdBOISU	stdBOIQL
5 yrs	0.7	0.6	0.95	0.6	0.9	0.95	0.679	0.795	0.6	0.6	0.95	0.89
20 yrs	0.7	0.6	0.4875	0.5	0.9	0.9875	0.79938	0.39875	0.5	0.3	0.4875	0.4475
40 yrs	0.7	0.6	0.74375	0.25	0.9	0.99375	0.52487	0.59938	0.25	0.45	0.74375	0.67375
80 yrs	0.7	0.6	0.74687	0.75	0.9	0.87187	0.78744	0.59969	0.75	0.45	0.74687	0.67438

Appendix 2: Results excluding respondents who misunderstood the DCE task

Appendix 2 Table 1: Summary of excluded respondents who misunderstood the DCE task and respondents in the remaining sample

	All respondents	Excluded respondents who arguably misunderstood	Remaining sample analysis	England *
n	3,669	1,422	2,247	
Mean age (s.d.)	46.5 (16.6)	48.6 (16.2)	45.1 (16.8)	NA
Age distribution				
18-40	39.9%	34.6%	43.2%	41.6%
41-65	42.1%	44.7%	40.5%	39.1%
Over 65	18.0%	20.7%	16.3%	19.3%
Female	54.3%	54.0%	54.5%	51.3%
Married/Partner	62.4%	64.8%	60.8%	NA
Employed or self-employed	47.3%	46.5%	47.9%	60.9%
Unemployed	6.2%	5.8%	6.4%	3.4%
Long-term sick	6.4%	7.3%	5.8%	5.3%
Full-time student	7.2%	4.9%	8.7%	7.3%
Retired	23.8%	26.7%	22.0%	13.5%
Secondary school is highest level of education	21.6%	26.4%	18.6%	
Degree or equivalent professional qualification	48.2%	47.0%	49.0%	
Health in general is very good or good	66.9%	63.7%	68.9%	
Limited by long term health	37.0%	41.4%	34.2%	
EQ-5D score, mean (s.d.)	0.78 (0.26)	0.76 (0.28)	0.80 (0.25)	0.86
Experienced serious illness in yourself	33.6%	36.2%	32.0%	
Experienced serious illness in family	74.5%	75.1%	74.2%	
Experienced serious illness in caring for others	33.5%	34.7%	32.7%	
Found DCE questions quite or very difficult to understand	7.6%	8.9%	6.8%	
Found attitudinal questions quite or very difficult to understand	6.6%	7.7%	5.9%	
Median completion time in minutes from consent to end of survey (Interquartile range)	21 (17-27)	20 (17-26)	22 (17-28)	

Notes: * Statistics for England in the Census 2001. Questions used in this study and the census are not identical. The census includes persons aged 16 and above whereas this study only surveys persons aged 18 and above. Age distribution is here reported as the percentage of all adults aged 18 and over.

Appendix 2 Table 2: Responses to practice questions for excluded respondents who misunderstood the DCE task and respondents in the remaining sample

Practice Question	Respondents	Practice question First response			Practice question Second response			Practice question Final response		
		n	A	B	n	A	B	n	A	B
1	All respondents	3,669	8.3%	91.7%	179	22.3%	77.7%	3,669	7.0%	93.0%
	Excluded respondents	1,422	10.1%	89.9%	72	19.4%	80.6%	1,422	8.7%	91.4%
	Remaining sample	2,247	7.1%	92.9%	107	24.3%	75.7%	2,247	6.0%	94.0%
2	All respondents	3,669	46.8%	53.2%	252	67.1%	32.9%	3,669	50.8%	49.3%
	Excluded respondents	1,422	27.9%	72.2%	88	55.7%	44.3%	1,422	30.6%	69.4%
	Remaining sample	2,247	58.7%	41.3%	164	73.2%	26.8%	2,247	63.5%	36.5%

Notes: Respondents who stated after the explanation of their choice that they did not still want to treat the same group were asked the question again, and this process was repeated allowing respondents up to 7 attempts in total before moving on to the next question.

Appendix 2 Table 3: HRQoL and life expectancy before and after treatment excluding respondents who misunderstood the DCE task

Variables	All	5 years	20 years	40 years	80 years
H	-0.051881***	-0.064297***	-0.052721***	-0.061813***	-0.051130***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
E	-0.136811***	-1.486658***	-0.411863***	-0.250984***	-0.050154***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HQ	0.057180***	0.069305***	0.058699***	0.060395***	0.060320***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ES	0.130886***	1.432319***	0.359007***	0.225150***	0.050150***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	-12107	-3244	-2090	-2536	-3160
Rho-squared	0.223	0.325	0.280	0.264	0.286
AIC	24222	6496	4188	5080	6327
BIC	24257	6526	4217	5109	6357
Mean VIF	2.10	2.13	1.85	5.61	1.76
stdH	-5.750397***	-6.429696***	-5.272089***	-6.181328***	-5.113010***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdE	-6.899348***	-7.433292***	-8.237261***	-10.039367***	-4.012287***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdHQ	6.252886***	6.930519***	5.869880***	6.039451***	6.031969***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdES	6.456234***	7.161597***	7.180147***	9.006004***	4.012018***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	44,940	13,860	8,380	9,940	12,760
Log likelihood	-11159	-3244	-2090	-2536	-3160
Rho-squared	0.284	0.325	0.280	0.264	0.286
AIC	22327	6496	4188	5080	6327
BIC	22362	6526	4217	5109	6357
Mean VIF	2.07	2.13	1.85	5.61	1.76
Observations	44,940	13,860	8,380	9,940	12,760

pval in parentheses; *** p<0.01, ** p<0.05, * p<0.1; H and E HRQoL and life expectancy before treatment; HQ and ES HRQoL and life expectancy after treatment

Appendix 2 Table 4: QALY gains and life expectancy before treatment/end of life excluding respondents who misunderstood the DCE task

Variables	All	5 years	20 years	40 years	80 years
QALY	0.378105***	4.199664***	1.024335***	0.574489***	0.232746***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
QALY_sq	-0.005354***	-0.853617***	-0.052240***	-0.017664***	-0.003016***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
E	-0.051728***	-0.430212***	-0.177651***	-0.089013***	-0.027102***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	-12670	-3266	-2049	-2616	-3252
Rho-squared	0.186	0.320	0.295	0.241	0.265
AIC	25347	6538	4103	5239	6509
BIC	25373	6560	4124	5260	6532
Mean VIF	6.27	7.00	5.08	5.22	8.02
stdQALY	20.232614***	20.998320***	20.486696***	22.979556***	18.619662***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdQALY_sq	-21.164541***	-21.340418***	-20.896071***	-28.262211***	-19.301296***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdE	-2.762355***	-2.151060***	-3.553029***	-3.560510***	-2.168166***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	-11254	-3266	-2049	-2616	-3252
Rho-squared	0.277	0.320	0.295	0.241	0.265
AIC	22514	6538	4103	5239	6509
BIC	22540	6560	4124	5260	6532
Mean VIF	5.93	7.00	5.08	5.22	8.02
QALY	0.307749***	3.296674***	0.914302***	0.492998***	0.187517***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
QALY_sq	-0.004281***	-0.640304***	-0.045515***	-0.017628***	-0.002328***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EOL_3mth	1.189960***	1.053927***	1.326816***	1.326196***	0.704581***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	-12480	-3254	-2004	-2708	-3358
Rho-squared	0.199	0.323	0.310	0.214	0.241
AIC	24966	6515	4014	5421	6723
BIC	24992	6537	4035	5443	6745
Mean VIF	5.42	7.85	4.99	5.13	7.42
stdQALY	16.354869***	16.483372***	18.286034***	19.719938***	15.001385***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdQALY_sq	-16.615309***	-16.007599***	-18.205976***	-28.204049***	-14.900992***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EOL_3mth	1.081820***	1.053927***	1.326816***	1.326196***	0.704581***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	-11402	-3254	-2004	-2708	-3358
Rho-squared	0.268	0.323	0.310	0.214	0.241
AIC	22811	6515	4014	5421	6723
BIC	22837	6537	4035	5443	6745
Mean VIF	5.62	7.85	4.99	5.13	7.42
Observations	44,940	13,860	8,380	9,940	12,760

pval in parentheses; *** p<0.01, ** p<0.05, * p<0.1; QALY – QALY gains from new treatment; QALY_sq – QALY squared
E –life expectancy before treatment; EOL – life expectancy before treatment ≤2 years and with survival gains≥ 3 months

Appendix 2 Table 5: QALY gains with life expectancy and HRQoL before treatment excluding respondents who misunderstood the DCE task

Variables	All	5 years	20 years	40 years	80 years
QALY	0.384801***	4.641321***	1.054246***	0.642400***	0.238122***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
QALY_sq	-0.005436***	-0.948974***	-0.053376***	-0.020285***	-0.003071***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
E	-0.052735***	-0.483926***	-0.181264***	-0.098784***	-0.027816***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdH	-0.296948***	-1.278359***	-0.578428***	-1.562883***	-0.429257***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	-12654	-3202	-2038	-2537	-3243
Rho-squared	0.188	0.333	0.298	0.264	0.267
AIC	25316	6413	4083	5082	6494
BIC	25351	6443	4112	5111	6524
Mean VIF	4.96	5.53	4.06	4.22	6.39
stdQALY	21.376507***	23.206606***	21.084918***	25.696004***	19.049750***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdQALY_sq	-22.277241***	-23.724347***	-21.350448***	-32.455529***	-19.651769***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdE	-2.932246***	-2.419631***	-3.625283***	-3.951376***	-2.225241***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
stdH	-0.885604***	-1.278359***	-0.578428***	-1.562883***	-0.429257***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	-11133	-3202	-2038	-2537	-3243
Rho-squared	0.285	0.333	0.298	0.264	0.267
AIC	22274	6413	4083	5082	6494
BIC	22308	6443	4112	5111	6524
Mean VIF	4.70	5.53	4.06	4.22	6.39
Observations	44,940	13,860	8,380	9,940	12,760

pval in parentheses; *** p<0.01, ** p<0.05, * p<0.1; QALY – QALY gains from new treatment; QALY_sq – QALY squared; H and E – HRQoL and life expectancy before treatment

Notes: stdH i.e. H/100 is used in the unstandardized QALY model to reflect the fact that H/100 is used to build the QALY. This has no effect on the direction of the coefficient; it just increases the magnitude by a factor of 100.

Appendix 2 Table 6: Summary of response to attitudinal questions for excluded respondents who misunderstood the DCE task and respondents in the remaining sample

	Normal life expectancy	**All respondents**	**Excluded respondents**	**Remaining sample**
Question	**Response**	**3,669**	**1,422**	**2,247**
	BOI			
1	The NHS should give priority to treating patients who are very ill	40.7%	34.2%	44.8%
	The NHS should give the same priority to treating all patients who are ill, regardless of how ill they are	59.3%	65.8%	55.2%
2	The NHS should give priority to treating patients who are very ill and will die early because of their illness	42.5%	35.2%	47.1%
	The NHS should give the same priority to treating all patients who are ill, regardless of how ill they are or when they will die	57.5%	64.8%	52.9%
3	The NHS should always give priority to treating patients who are very ill and will die early because of their illness, even if they only get a small amount of benefit from treatment	9.5%	9.4%	9.7%
	The NHS should give priority to treating patients who are very ill and will die early because of their illness, but only if they get a large amount of benefit from treatment	44.4%	38.3%	48.3%
	The NHS should give the same priority to treating all patients, regardless of how ill they are or when they will die	46.1%	52.4%	42.0%
8	The NHS should give priority to treating patients who are very ill and will die early because of their illness	10.6%	8.4%	12.0%
	The NHS should give priority to treating patients who will get the largest amount of benefit from treatment	44.6%	41.4%	46.7%
	The NHS should give the same priority to treating all patients	44.8%	50.2%	41.3%
	EOL			
4	The NHS should give priority to extending the life of patients who are expected to die soon, even if this is the natural end of their life	6.3%	7.5%	5.5%
	The NHS should give priority to patients expected to die soon, but only if it means they die before the natural end of their life	38.4%	30.7%	43.3%
	The NHS should give the same priority to treating all patients, regardless of how ill they are or when they will die	55.3%	61.8%	51.2%
5	The NHS should give priority to extending the life of patients who are expected to die soon, even if this means they live in very poor health	3.9%	4.1%	3.7%
	The NHS should give priority to extending the life of patients who are expected to die soon, but only if they would live in a reasonable level of health	56.4%	50.2%	60.3%
	The NHS should give the same priority to treating all patients, regardless of how ill they are or when they will die	39.7%	45.7%	36.0%

	Normal life expectancy	All respondents	Excluded respondents	Remaining sample
Question	**Response**	**3,669**	**1,422**	**2,247**
7	The NHS should give priority to extending the life of patients expected to die soon	12.0%	10.8%	12.8%
	The NHS should give priority to treating patients who will get the largest amount of benefit from treatment	88.0%	89.2%	87.2%
	Therapeutic Improvement			
6	The NHS should give priority to treatments that give a large amount of benefit to a small number of patients	8.1%	8.9%	7.7%
	The NHS should give priority to treatments that give a small amount of benefit to a large number of patients	9.5%	11.3%	8.3%
	The NHS should consider the amount of benefit a treatment gives overall, rather than considering how it is shared out among different numbers of patients	82.4%	79.8%	84.0%
	Combined			
9	The NHS should give priority to treating patients who are very ill and will die early because of their illness	13.1%	11.0%	14.4%
	The NHS should give priority to treating patients who will get the largest amount of benefit from treatment	52.1%	48.5%	54.4%
	The NHS should give priority to treating patients who will live for a long time and be in good health after treatment	34.8%	40.5%	31.2%

Appendix 3: Analysis on attitudinal questions and DCE responses

Appendix 3 Table 1: Test attitude question 2 [Burden 42% vs. All the same 58%]

	Variables	All	5yrs	20yrs	40yrs	80yrs
Prefer BOI	QALY	0.263849***	3.750114***	0.798500***	0.386393***	0.166535***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	QALY_sq	-0.003709***	-0.717338***	-0.041634***	-0.013412***	-0.002048***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	BOI	0.036957***	0.350807***	0.059065***	0.055351***	0.021540***
		(0.000)	(0.000)	(0.002)	(0.000)	(0.000)
	Observations	31,160	8,600	6,320	7,700	8,540
	Log likelihood	-9356	-2163	-1680	-2300	-2375
	Rho-squared	0.134	0.274	0.233	0.138	0.198
Same Priority regardless of BOI	QALY	0.288450***	3.591170***	0.720589***	0.420029***	0.177417***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	QALY_sq	-0.003964***	-0.707044***	-0.034468***	-0.014630***	-0.002087***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	BOI	0.000810	-0.052535	-0.040132***	0.025085***	-0.010099***
		(0.731)	(0.196)	(0.010)	(0.000)	(0.000)
	Observations	42,220	11,840	8,880	10,080	11,420
	Log likelihood	-12343	-2975	-2353	-2932	-2996
	Rho-squared	0.156	0.275	0.235	0.161	0.243

pval in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Question 2 : The NHS should	All	5	20	40	80
give priority to treating patients who are very ill and will die early because of their illness	42.5%	42.1%	41.6%	43.3%	42.8%
give the same priority to treating all patients who are ill, regardless of how ill they are or when they will die	57.5%	57.9%	58.4%	56.7%	57.2%

Appendix 3 Table 2: Test attitude question 5 (Preference for those at EOL 60% vs. Same priority 40%)

	Variables	All	5yrs	20yrs	40yrs	80yrs
Prefer to treat those at EOL	QALY	0.289660***	3.465257***	0.790120***	0.432395***	0.173923***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	QALY_sq	-0.004004***	-0.645310***	-0.038083***	-0.014956***	-0.002062***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	EOL	0.713568***	0.746571***	0.419514***	0.733828***	0.453378***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	Observations	44,220	12,120	9,320	10,860	11,920
	Log likelihood	-12780	-2920	-2425	-3109	-3216
	Rho-squared	0.166	0.305	0.249	0.174	0.221
Same priority for all	QALY	0.269786***	2.937967***	0.718618***	0.359284***	0.177529***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	QALY_sq	-0.003779***	-0.546943***	-0.034800***	-0.012604***	-0.002207***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	EOL	0.461005***	0.432922***	0.308074***	0.355861***	0.104242
		(0.000)	(0.000)	(0.000)	(0.000)	(0.154)
	Observations	29,160	8,320	5,880	6,920	8,040
	Log likelihood	-8613	-2170	-1581	-2081	-2171
	Rho-squared	0.148	0.248	0.224	0.132	0.221

pval in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Question 5:The NHS should:	All	5	20	40	80
give priority to extending the life of patients who are expected to die soon	60.3%	59.3%	61.3%	61.1%	59.7%
give the same priority to treating all patients, regardless of how ill they are or when they will die	39.7%	40.7%	38.7%	38.9%	40.3%

Appendix 3 Table 3: Test attitude question 6 (Therapeutic improvement: Concentration 8% vs. dispersion 9% vs. overall QALY gain 82%)

	Variables	All	5yrs	20yrs	40yrs	80yrs
Large amount benefit to small number of Patients (concentration)	QALY	0.222548***	4.428066***	0.608634***	0.369315***	0.131028***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	QALY_sq	-0.003153***	-0.960060***	-0.032567***	-0.013672***	-0.001584***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	BOI	0.015528***	0.246911**	-0.028703	0.022020*	0.008351
		(0.009)	(0.030)	(0.449)	(0.072)	(0.204)
	Observations	5,980	1,800	1,240	1,580	1,360
	Log likelihood	-1867	-447.0	-357.5	-480.3	-403.9
	Rho-squared	0.0992	0.283	0.168	0.123	0.143
Small amount benefit to a large number of patients (dispersion)	QALY	0.177368***	1.700498***	0.555886***	0.257769***	0.119564***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	QALY_sq	-0.002534***	-0.304565***	-0.026015***	-0.009224***	-0.001518***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	BOI	0.003366	0.063837	-0.048391	0.046921***	-0.010773*
		(0.491)	(0.487)	(0.143)	(0.000)	(0.055)
	Observations	6,960	1,700	1,720	1,880	1,660
	Log likelihood	-2224	-528.2	-495.1	-605.7	-499.1
	Rho-squared	0.0779	0.104	0.169	0.0704	0.132
QALY benefit Overall	QALY	0.297793***	3.875474***	0.803054***	0.431632***	0.181477***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	QALY_sq	-0.004107***	-0.745009***	-0.039694***	-0.014765***	-0.002145***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	BOI	0.019981***	0.122691***	0.013221	0.039859***	0.006399***
		(0.000)	(0.000)	(0.333)	(0.000)	(0.002)
	Observations	60,440	16,940	12,240	14,320	16,940
	Log likelihood	-17626	-4138	-3178	-4137	-4491
	Rho-squared	0.159	0.295	0.251	0.166	0.235

pval in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Question 6 - The NHS should:	All	5	20	40	80
give priority to treatments that give a large amount of benefit to a small number of patients	8.1%	8.8%	8.2%	8.9%	6.8%
give priority to treatments that give a small amount of benefit to a large number of patients	9.5%	8.3%	11.3%	10.6%	8.3%
consider the amount of benefit a treatment gives overall, rather than considering how it is shared out among different nos. of patients	82.4%	82.9%	80.5%	80.5%	84.9%