EEPRU

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

Research Report

Title: Update: Eliciting societal preferences for weighting QALYs according to burden of illness, size of gain and end of life

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

Correspondence to: Donna Rowen, HEDS, ScHARR, Regent Court, 30 Regent Street, Sheffield, S1 4DA.
Email: d.rowen@sheffield.ac.uk

Number RR018
Date March 2014

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 then 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.

Acknowledgements

We are grateful to the survey respondents. We are grateful to Gavin Roberts and Danny Palnoch from DH for their comments throughout this project and would also like to thank Angela Robinson for her input into the survey framework. We would also like to thank Emily Lancsar, Rachel Baker, Cam Donaldson, Jose Luis Pinto Prades and Richard Cookson for comments on earlier versions. Only the authors are responsible for any remaining errors.

EXECUTIVE SUMMARY

This report presents the findings of a study to elicit societal preferences and weights for Quality Adjusted Life Year (QALY) gains across three characteristics: 1) Burden of illness (BOI) from a medical condition given current health care interventions – defined as the QALY loss per patient from a condition due to premature mortality and/or reduced health-related quality of life (HRQOL) measured against life expectancy and health-related quality of life without the condition; 2) Therapeutic improvement (TI) - whether preferences for large QALY gains are disproportionately larger than the size of the gain (e.g. weight a QALY gain of 2 more than 4 times a QALY gain of 0.5) and 3) End of life (EOL) - defined by NICE as expected survival of less than 2 years and expected survival gain of 3 months or more.

Methods

A survey using a Discrete Choice Experiment (DCE) was conducted with an online general population sample using an existing panel. Respondents were asked to choose whether they thought the NHS should treat patient group A or B, who differed in terms of four attributes: life expectancy without treatment, HRQOL without treatment, survival gain from treatment and HRQOL gain from treatment. These attributes were used to derive BOI, QALY gain and EOL. The questionnaire had four variants, each with a different life expectancy without the condition (5, 20, 40 and 80 years). Each respondent answered questions for one variant and made comparisons between groups with the same life expectancy without the condition. Choices were analysed using conditional logistic regression with a range of specifications. Robustness across the four levels of life expectancies without the condition and to various exclusions was examined. Weights were estimated using the marginal rate of substitution.

Results

In total, 3669 respondents completed the survey. The sample was largely representative of the population of England for age and gender, but there were some differences in other characteristics. Regression results indicated that respondents preferred to treat patients with larger QALY gains, but at a diminishing rate meaning there was no support for TI. Respondents preferred to treat patients with a shorter life expectancy (EOL). Results suggested some support for BOI but were not robust across alternative model specifications. The coefficients varied as life expectancy without the condition varied. Regressions estimated excluding respondents who were identified as possibly misunderstanding the DCE task (remaining sample of 2247 respondents) had positive, significant and robust coefficients for BOI. Using the marginal rate of substitution to estimate weights indicated that 1 unit of

BOI is equivalent to 0.04 QALYs gained, and EOL is equivalent to 3.331 QALYs gained (assuming that the value of a QALY does not change with size of QALY gain).

Discussion

This study provides the first attempt to operationalize the concept of BOI by combining the conventional notion of severity (in terms of poor health) with survival, using QALY loss attributable to the condition. The results indicate general support for maximising QALY gains, but at a diminishing rate, meaning that the evidence did not support the idea of TI. The results indicate some support for BOI as a consideration when weighting QALYs. There is robust and consistent support for EOL in general (but this conceptually overlaps with BOI and the two should not be used together). Overall the results indicate that a QALY is not a QALY regardless of the burden of the disease or life expectancy and provide a basis for determining appropriate QALY weights.

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. (1) The approach is designed to improve the efficiency of health care spending and typically assumes that an additional QALY is worth the same regardless of who gets it. However, agencies that use cost per QALY analyses allow for other considerations explicitly or have been shown to do so implicitly in their decision making. (2)It has long been recognised that the QALY approach can incorporate a more complex algorithm than simply assuming 'a QALY is a QALY'. (3) At the same time there is emerging evidence that members of the public weight some QALY gains more highly than others depending on who receives them. (4-7)

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 fair innings. For an attribute to be used in cost per weighted QALY analysis it needs both to be supported by normative argument and for empirical evidence to quantify its size. The empirical evidence can be elicited in surveys of the general public on the grounds that they are potential tax payers or the basis of democratic principles. 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 additional QALYs are of equal value. There is evidence for a preference for those in worse health, though this is not found in all studies.(7) There is also evidence that tends to favour younger recipients and those who are not responsible for their condition.(8) The legal basis for some of these attributes can be challenged, so for example the use of age may be regarded as discriminatory (as would be the case in the UK) and responsibility can be hard to establish for most medical conditions. Therefore 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.

An important consideration is the way severity is defined and measured. The earlier literature tended to focus on severity in terms of the health related quality of life of the recipient before treatment (4;9) finding that respondents often gave gains at the lower end of the 1-0 full health-dead scale a higher weight than gains at the higher 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, (10)

“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’’. 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.” (10)(Hansson et al 1994, p353). It has been argued further that equity considerations would take into account a person’s whole health profile and not simply that from today. This is the basis of a fair innings criterion, whereby the weight of a QALY gain 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.(11) However, decisions are made for the future and it could be argued that prospective health should be the focus for decision making.(4)

The research reported in this paper attempts to operationalize 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) .(1) In a consultation document, the English Department of Health set out a new mechanism for pricing drugs in the UK known as Value Based Pricing.(12) Value-Based Pricing aims to assess the cost-effectiveness of medicines taking into account a broader scope of value, including the severity of disease and wider societal benefits. The consultation specifies severity in the same way as Hansson and Colleagues (1994) and refers to this as burden of illness (BOI). BOI is measured using the outstanding QALY loss suffered by patients with current treatments relative to their prospects in the absence of the disease.

The consultation document proposes 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 attribute 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,(13) but there is some evidence to suggest that there may be a threshold below which respondents prefer to concentrate QALY gains. One study found that respondents tended to prefer to concentrate life year gains below a threshold of 9 years (14) and another study found that there was a threshold at 2.6 years. (8) The precise threshold is likely to depend on the context, but the existing evidence provides some support for concentration, and thus TI, but in a restricted range.

In addition, this work was being undertaken in the UK policy environment where there already exists another attribute that is used in HTA appraisal to weight QALYs. This is the

'end of life' (EOL) 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 2 years and a survival gain of 3 months or more (provided the condition is a 'rare' disease).(15) This attribute has also been included in this research.

The aim of the research presented in this paper was to elicit societal preferences for the following attributes: 1) BOI from a medical condition given current health care interventions – defined as the QALY loss per patient from a condition due to premature mortality and/or reduced health-related quality of life measured against life expectancy and health-related quality of life without the condition; 2) TI - whether preferences for large QALY gains are disproportionately larger than the size of the gain (e.g. weight a QALY gain of 2 more than 4 times a QALY gain of 0.5) and 3) EOL - defined by NICE as expected survival of less than 2 years and expected survival gain of 3 months or more. This paper presents the methods developed for operationalizing these attributes, the survey to elicit the preferences of the general public using a discrete choice experiment, regression analyses of the survey data and QALY weights. The results are detailed together with a discussion of their implications for research in this field and health care policy.

Methods

The framework

The components of the attributes of BOI, TI and EOL are shown in Figure 1. These components are 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. At that point they have a health profile without treatment, which for simplicity is represented by health-related quality of life or health H and life expectancy E. To estimate the BOI in terms of the QALY loss associated with the condition it is necessary to establish an expected or target level of health and life expectancy. In this study the expected profile without the condition is assumed to be 100% health with life expectancy N. The improvement from treatment is represented by a gain in health Q and an improvement in survival S. BOI is the loss of health and life expectancy from their expected or target levels, measured as QALY loss from morbidity (areas B+D in Figure 1), and QALY loss from premature mortality (areas A+C in Figure 1), generated as 100*N - area F. QALY gain is areas D+C, and end of life is where E is 2 years or less and S is 3 months or more. A diagram like this was used to present the different combinations of attributes to respondents in the survey.

Elicitation technique

A DCE based on pairwise comparisons was chosen as the method to elicit preferences. It permits the simultaneous consideration of different attributes in a format that is amenable to being administered online and has been successfully employed by Lancsar and colleagues in a survey eliciting QALY weights.(16) Several preparatory studies were undertaken to determine the choice of DCE, the online mode of administration and question framing. These included: a review on the social value of a QALY; a large preparatory online survey to pilot DCE and person trade-off questions designed to elicit preferences for BOI, QALY gain and EOL; a qualitative survey to further explain the findings of the large preparatory online survey; and a six-arm online and face-to-face survey examining different framings of questions and mode of administration to determine which was most appropriate for eliciting preferences for BOI, QALY gain and EOL. For further details see Brazier et al (2014).(17)

Selection of attributes and levels

The valuation survey consisted of a pairwise comparison DCE with 4 attributes: life expectancy without treatment (E), survival gain from treatment (S), health before treatment (H), health gain from treatment (Q) (Figure 1). There were 4 different DCE designs, each with a different level of life expectancy without the condition (N): 5 years, 20 years, 40 years or 80 years resulting in 4 variants of the questionnaire. Respondents saw one questionnaire variant i.e. one level of life expectancy without the condition across all DCE pairs that they attempted. This was due to concerns that different levels of life expectancies without the condition would be confusing for respondents and would highlight the differences in age of the two profiles, where age is a consideration that is not politically desirable and is therefore not a consideration in the framework. The levels of each of the attributes for each level of life expectancy without the condition are outlined in Table 1. These were selected to cover a full range of potential levels, including patient groups involving children, but also to ensure precision over the more common characteristics of interventions in the UK where patients have a small number of years of life expectancy remaining without the condition, with small QALY gains.

DCE design

A full factorial design using the attributes and levels specified in Table 1 would result in a very 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 (18;19) and the true model specified in such a way as to allow for the estimation of an additive model including all parameters of interest (using derived variables for QALY gain, BOI and EOL rather than the attributes H, Q, E and S in the design). Each experiment was designed

to minimise the amount of correlation between the derived variables. Impossible profiles (such as profiles involving health after treatment 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 'card blocs') of 10 pairs. Each combination contained pairs for one level of life expectancy without the condition.

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 in variables considered in this survey with QALY gains from 0.005 up to 63, BOI from 1 to 80 QALYs lost, and life expectancy from 0.25 to 60 years.

Analysis of data

The DCE data was modelled based on a random utility theory (RUT) framework. Within the RUT framework, utility U_{ij} for an individual *i* was 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 had within a choice set. The alternative chosen by the individual was assumed to confer greater utility than any other alternative. Choices were based on a set of attributes captured in V_{ij} and other influencing factors that were not observed were captured by the random component. DCE data provide the alternatives that individuals chose, in this case whether respondents thought the NHS should treat patient group A or patient group B. These were modelled using the conditional logistic model which models the probability that individual *i* chose profile *j = A, B,* for example, so the probability of an individual choosing to treat patient group A over B was 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 derived from choosing to treat patient group A and B, respectively. There were multiple observations for each individual and the estimated models cluster the standard errors at the respondent level to allow for respondent effects.

V was modelled as a function of attributes *z*:

$$V = f\ (z) \tag{3}$$

Where *z* represents:

- *BOI* representing burden of illness from both premature death (BOISU, A+C in Figure 1) and health loss (BOIQL, B+D in Figure 1) generated using $N - \frac{H}{100}E$,
- *QALY* representing QALY gain from survival (C in Figure 1) and QALY gain from improved health (D in Figure 1) generated using $S\left(\frac{H+Q}{100}\right) + \frac{Q}{100}E$, and
- a dummy variable to represent *EOL* using the NICE definition of 2 years life expectancy or less (E≤2 years) and survival gain of 3 months or more (S≥3 months)

BOI and EOL were not included in the same model specification due to conceptual overlap in these variables: as EOL profiles have life expectancy of 2 years or less these will also have a large BOI. Models were estimated that split BOI into the components of BOI from premature death (BOISU) and health loss (BOIQL) to determine whether these components differ in their impact on utility.

Model specification

The survey was designed to estimate an additive model where each attribute was entered as an independent main effect:

$$V = \beta_{ij}\mathbf{z}_{ij}\ + \gamma_{ij}\mathbf{w}_{ij}\ \ + \varepsilon \tag{4}$$

where *V* represents utility, **z** represents a vector containing the variables described in the section above and **w** represents the squared terms of each. This additive model specification was chosen to keep the model as simple and transparent as possible.

The simple model for QALY gain and BOI is:

$$V_{(1)}^{BOI} = \beta_1 QALY + \beta_2 BOI + \varepsilon \tag{5}$$

A more complex regression model including a QALY gain squared term to account for TI is:

$$V_{(2)}^{BOI} = \beta_1 QALY + \beta_2 BOI + \beta_3 QALY^2 + \varepsilon \tag{6}$$

where a positive value for β_3 indicates an increasing marginal utility as QALY gains increase. Plots indicated that a QALY cubed term was not appropriate for the distribution of the data.

Different model specifications are reported representing models examining: BOI and QALY gain; BOI (model $V_{(1)}^{BOI}$), QALY gain and TI (model $V_{(2)}^{BOI}$); BOI split into survival and health, QALY gain and TI (model $V_{(4)}^{BOI}$); EOL and QALY gain (model $V_{(1)}^{EOL}$); and EOL, QALY gain and TI (model $V_{(2)}^{EOL}$).

Model performance

Performance of all regression models was assessed using the log-likelihood, Rho-squared, Akaike Information Criterion (AIC) (20) and the Schwarz Bayesian Information Criterion (BIC). (21) Models were preferred with higher log likelihood, larger Rho-squared and lower AIC and BIC.

Robustness of results

Robustness of results was assessed for the impact of excluding responses from individuals who may have not understood or engaged with the survey. A number of exclusion criteria were examined to identify these individuals including: those who reported that they found the survey quite or very difficult, those who took less than 5 minutes or more than 60 minutes to complete the survey and those who selected to treat the same patient group for all 10 questions (this may indicate respondents were selecting either all left or all right sides of the screen). In addition respondents were excluded if they were identified as having possibly misunderstood the DCE task. These were respondents who chose to treat the patient group with a larger number of total lifetime QALYs after treatment, but smaller QALY gain from treatment and lower BOI before treatment than the other patient group. These respondents were excluded on the grounds that they seemed to choose the profile they thought was best for them personally or that they wanted to live in, not the profile which was best from a societal perspective where the patients were most “deserving” of treatment.

A final check was done by separately excluding the first and last survey questions that respondents completed on the basis that these responses may be less reliable. Questions were allocated to respondents in a random order and therefore exclusion of the first or last question should have no systematic impact on results.

Estimating weights

The marginal rate of substitution (MRS) was used to indicate the value for BOI in terms of QALY gain. The MRS was estimated using the ratio of the marginal utilities:

$$MRS^{BOI} = -MU_{BOI}/MU_{QALY} = -\frac{\partial U}{\partial BOI}/\frac{\partial U}{\partial QALY} \quad (7)$$

Where MU_{BOI} represents the marginal utility of BOI and MU_{QALY} represents the marginal utility of the QALY gain, generated using the first order partial derivative of the utility function with respect to BOI and QALY gain respectively. For the model specified in equation (5) this is:

$$MRS_{(1)}^{BOI} = -\frac{\widehat{\beta}_2}{\widehat{\beta}_1} \quad (8)$$

For the model specified in equation (6) this becomes:

$$MRS_{(2)}^{BOI} = -\frac{\widehat{\beta}_2}{\widehat{\beta}_1 + 2\widehat{\beta}_3 QALY} \quad (9)$$

MRS for EOL, MRS^{EOL}, is generated using equivalent regression specifications involving EOL rather than BOI. The regressions selected to generate the coefficients were estimated using all data collected in the survey, and used observations across all variants of the questionnaire (5, 20, 40 and 80 years of life expectancy without the condition). The rationale for using all data was to obtain a representation of the data across all levels of life expectancy without the condition. The standard error (S.E.) and the 95% confidence interval (95% CI) of the MRS were calculated using the Delta method (see e.g. Hole, 2007).(22)

The survey

Respondents from an online panel were contacted via email to participate in the survey. Respondents were sampled to be representative of the UK adult population in terms of age (minimum age 18) and gender. At the start of the survey respondents read an information page and gave informed consent to participate in the survey. Respondents were then shown a short video explaining the questions. 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 which involved a “feedback screen” including an explanation of their choice with a chance for respondents to change their mind. The first question had one “dominant” alternative, which had a larger QALY gain than the other alternative, while all other attributes were the same. The second question also had one dominant alternative, which had a larger BOI, while all other attributes were the same. Figure 2 shows the information displayed on the first screen of practice question 1 with a life expectancy without the condition of 20 years. Figure 3 shows the feedback screen for a respondent who chose to treat Patient group A. Respondents were asked on the feedback screen whether they still wished to treat that Patient group. They started the next question if they did not change their mind, and were shown the question from the first screen again if they did change their mind. 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 completed 10 DCE questions, 9 questions on attitudes, and 17 questions covering EQ-5D of their own health, socio-demographics, difficulty of understanding of the DCE and attitudinal questions, and what they thought of the survey. The ordering of the 10 DCE questions was random for each respondent. Attitude questions were included to determine respondents’ general views on BOI, TI and EOL. This enables interpretation of the results of the practice questions, regression modelling and weightings, as these should be in accordance with the results of the attitude questions that remove the complexities and intricacies of the DCE questions. The framing of the DCE questions, the attitude questions and the socio-demographic and understanding questions were all piloted in the preparatory studies undertaken prior to this survey.

The Data

A total of 3669 respondents completed the online survey, providing a response rate of 55% of people who accessed the survey. All respondents completed every question. No respondents were excluded from the main analysis. Characteristics of the sample were compared to the general population in England in Table 3. In comparison to the general population of England, the sample was largely representative for age and gender, but had higher proportions of individuals who were unemployed, long-term sick and retired, and lower proportions of individuals who were employed or self-employed. The sample also had a lower EQ-5D score (23) 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 a serious illness in themselves. A large proportion of the sample, 48.2%, had a degree or equivalent professional qualification.

RESULTS

Completion times

Median completion time 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 a small proportion (≈ 5%) of respondents took over an hour. A large proportion of respondents (65%) spent between 7 and 10 minutes on watching the introduction video and completing the practice questions, suggesting the respondent watched the video (which lasted approximately 5 minutes 50 seconds) and then considered the practice questions. However, some respondents had long times (up to 1 hour) that suggested they may have left the survey idle in this time and therefore there are doubts as to whether these respondents watched the video.

Practice questions

In practice question 1, respondents overwhelmingly (93.0%) chose to treat the group with the highest QALY gain, all other things the same, and this was consistent across the different variants, varying from 90.7% to 92.5% (see Table 4). In practice question 2 there was little evidence (50.8%) that, other things equal, respondents preferred to treat the patient group with higher BOI, with 46.8%, 54.3%, 52.3% and 50.7% of respondents across the four questionnaire variants choosing this group. Following the feedback on their choice, where respondents were offered the option to change their mind up to 7 times in each practice question, the final responses differed to their first response for 2-4% of respondents. These results suggest that the feedback did clarify the profiles for some respondents and that a small number of respondents altered their choices following the feedback. In robustness analyses, excluding respondents identified as possibly misunderstanding the DCE task increased the proportion choosing to treat the patient group with higher BOI to 63.5% (remaining n=2247).

Regression results

Table 5 presents the main regression results. The coefficients varied across the different variants meaning that the coefficients vary as life expectancy without the condition varies.

Across all models, QALY gain had a positive and significant coefficient, indicating that respondents preferred profiles with higher QALY gains. The coefficient for the QALY gain squared term, when included, was negative and significant across all variants and models, indicating that QALY gains were preferred at a decreasing rate. This means that there was no support for TI; in fact the opposite was observed.

The coefficient for BOI, when included, was small, positive and significant for the pooled analysis, indicating that respondents preferred profiles with higher BOI. However, results by life expectancy variant varied, where BOI was non-significant for the 20 year variant in models $V_{(1)}^{BOI}$ and $V_{(2)}^{BOI}$; and for the 80 year variant in model $V_{(1)}^{BOI}$. BOI squared was tested but did not improve the models and although it was statistically significant in some models, the BOI main effects term was no longer statistically significant hence the squared term has not been included here. In models where BOI was split into health loss and life expectancy loss, coefficients were not always significant across variants, but when they were, BOI from health loss (BOIQL) was negative while burden from life expectancy loss (BOISU) was positive. This indicates that respondents were more likely to choose to treat patient groups with higher burden from life expectancy and lower burden in health. These coefficients were not consistent in direction between the questionnaire variants, but the coefficients were not statistically significant where the direction was inconsistent.

The coefficient for EOL was positive and significant indicating respondents gave greater weight to shorter life expectancy before treatment when survival gains were greater than 3 months.

Comparison of model performance

Model performance was improved by the inclusion of the QALY squared term. The best performing models using AIC, BIC, log likelihood and Rho-squared were the specifications with either EOL or BOI split into health and life expectancy losses.

Robustness of results

The robustness of results was examined as 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 each of 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; 23 individuals who chose the same option for all their DCE questions; all responses to the first DCE non-practice questions; and all responses to the last DCE non-practice questions. The exclusions impacted on the magnitude of the coefficients but not their sign or significance with the exception of BOI. Excluding individuals who took less than 5 minutes or more than 60 minutes to complete the survey changed the significance for BOI when life expectancy

without the condition was 80 years. This means that, contrary to expectations, the BOI coefficient was negative and significant in model $V_{(1)}^{BOI}$ when life expectancy without the condition was 80 years. Excluding the first question for each individual meant that the BOI coefficient became insignificant in model $V_{(1)}^{BOI}$ when life expectancy without the condition was 5 years. Excluding the last question for every individual meant that the BOIQL coefficient became insignificant in model $V_{(3)}^{BOI}$ when life expectancy without the condition was 40 years.

Second, 1422 respondents who were identified as possibly having misunderstood the DCE task were excluded. For questionnaire variants with 5, 40 and 80 years life expectancy without the condition, some respondents did not answer a question that could be used to implement this exclusion criterion. These 369 respondents remain in the analysis although it is possible that some of these respondents may have also misunderstood the DCE task. In regressions estimated on the remaining sample (n=2247) there was no impact on the coefficients for QALY gain, QALY gain squared and EOL in terms of significance and direction although there were some changes in magnitude. However, the coefficient for BOI was larger, positive and significant for all models. When BOI was split between losses in life expectancy and health the coefficients were always positive and were significant with the exception of the 80 year variant for BOI from health loss (BOIQL).

Weights

The $MRS_{(1)}^{BOI}$ of 1 unit loss of BOI is -0.040 QALYs (95% CI (-0.068, -0.013)). In other words, this indicates that if BOI increases by 1 unit, the level of utility is maintained by a QALY loss of 0.040 QALYs. This calculation assumes that the social value of a QALY gain does not change with size of QALY gain. If the QALY gain squared term is included in the model the $MRS_{(2)}^{BOI}$ is -0.064 QALYs when the QALY gain is 1 (95% CI (-0.082, -0.047)). However, the weighting for one extra unit of BOI now differs depending on the size of QALY gain as shown in Table 6, ranging from -0.063 to -0.141 as QALY gain changes from 0.05 to 20.

The $MRS_{(1)}^{EOL}$ of moving from not being EOL to being EOL is -3.331 QALYs (95% CI (-3.711, -2.950)). In other words, this indicates that by moving from not being EOL to being EOL, the level of utility is maintained by a QALY loss of 3.331. If the QALY gain squared term is included in the model the $MRS_{(2)}^{EOL}$ is -2.229 QALYs for a QALY gain of 1 (95% CI (-2.438, -

2.020)). Allowing the value of a QALY gain to vary by the size of the QALY gain results in a range in $MRS_{(2)}^{EOL}$ of -2.170 to -4.875 as QALY gain changes from 0.05 to 20.

Attitudinal questions

The results suggest some support for BOI, with between 40.7% and 53.9% of respondents agreeing that the NHS should give priority to treating patients with BOI over giving the same priority to treating all patients regardless of how ill they are or when they will die. However, only 9.5% of respondents agreed that the NHS should give priority to treating patients with BOI if the patients only get a small amount of benefit from treatment.

There was some support for EOL, with between 44.7% and 60.3% of respondents agreeing that the NHS should give priority to patients who are expected to die soon over giving the same priority to treating all patients regardless of how ill they are or when they will die. However, only 13.1% of respondents agreed that the NHS should give priority to patients who are expected to die soon if they were at the natural end of their life, and only 3.9% agreed with giving priority to these patients if they live in very poor health. Furthermore only 12.0% of respondents agreed that the NHS should give priority to extending the life of patients expected to die soon over giving the same priority to all patients.

There was very little support for TI and concentrating gains, with only 8.1% of respondents agreeing that the NHS should give priority to treatments giving a large amount of benefit to a small number of patients.

Overall the responses to the attitudinal questions indicated that most respondents believed that the NHS should give preference to the group with the largest treatment gain over BOI or EOL (see Table 7). A large proportion of respondents consistently indicated that the same priority should be given to all patients implying that they did not want size of QALY gain, BOI or EOL to be taken into account, and this was the modal response.

Respondent views of the survey

The majority of respondents, 77.7%, reported that the DCE questions were either very easy or fairly easy to understand, varying from 76.2% to 79.5% across the different life expectancies without the condition. The majority of respondents, 77.7% also reported that the attitudinal questions were either very easy or fairly easy to understand, varying from 76.2% to 80% across the life expectancies without the condition.

DISCUSSION

Summary

This was the first study to examine societal preferences for BOI alongside TI and EOL. It was a large DCE survey using an existing online panel drawn from the general population. Respondents preferred to treat patients who had larger QALY gains but this was at a diminishing rate. They also preferred to treat patients at the EOL using the NICE definition. The results for BOI were less robust across variants of the questionnaire, but suggested some modest support for BOI. Using the MRS to estimate weights indicated that 1 unit of BOI is equivalent to 0.04 QALYs gained, and EOL is equivalent to 3.331 QALYs gained. Attitudinal questions seemed to support the regression results for QALY gains and BOI although less so for EOL.

QALY gains and therapeutic improvement

The results of this survey indicate that respondents tend to choose to treat the group with the larger QALY gain, but they do not support the notion of TI set out in the VBP consultation document.(12) Although not directly comparable in terms of the attributes included, the Lancsar et al (2011) study undertaken in the UK found QALY gains to have a positive and statistically significant impact but again at a declining rate, and so did a recent study in Australia by Norman et al (2013). This is also consistent with studies finding evidence for dispersing life year gains above a threshold,(8;14) although the current survey was not designed to examine this.

EOL

The regression results showed support for EOL across the regression models, with evidence for a preference to treat those who were at the end of their life. However, the responses to the attitudinal questions cast some doubt on the strength of this finding as this was not a view held by the majority. The evidence of a preference for EOL is consistent with a small survey conducted by Shah et al (2013) which found weak support for EOL. Conversely, their larger follow-on study indicated little support for EOL,(24) and a survey by Linley and Hughes (2013) also found no support for EOL. Whilst these contradictory results are surprising, this may in part be due to differences in framing across the surveys.

BOI

The findings from the attitude survey and the DCE suggest some support for BOI, though the findings are not consistent across all models. Whilst the responses to practice question 2 found little support for BOI with 50.8% of respondents choosing the option with higher BOI, there were other questions in the survey with dominant pairs or near dominant pairs in terms

of BOI (34 pairs) that provided more support. Analyses of these revealed a majority of respondents typically choosing the group with the higher BOI at 52-85%.

Splitting the BOI term indicated different effects of burden from health loss (BOIQL) and burden from shorter life expectancy (BOISU). The surprising negative coefficients on BOIQL were probably observed because BOIQL was not solely attributable to health loss (in terms of Figure 1, it is the product of without treatment health H and life expectancy E). BOIQL can increase due to either a reduction in H or an increase in E, and these are likely to impact differently on how the without treatment profile is regarded. Therefore BOIQL cannot be seen as a test of the conventional severity argument and BOI should not be split into a health effect and a survival effect, since by definition BOI is composed of both. However, some recent large scale studies with members of the general population, including Lancsar et al (2011) and Norman et al (2013), found that respondents were less likely to choose to treat patients with a lower quality of life before treatment, H, which is consistent with our findings.

BOI provides a broader notion of severity than previous research since it incorporates the impact on health and life expectancy over the patients' future life. Furthermore the survey incorporates end of life. On theoretical grounds BOI and EOL should not be used together either in regressions or to weight QALY gains as profiles with EOL will also have a large BOI. There is an important policy decision about which measure to include. The advantage of BOI is that it incorporates a number of different equity concerns. A further attraction of the way BOI has been operationalized is that it is measured using QALYs and so is compatible with cost-effectiveness in terms of cost per QALY gain.

Weights

The weights for BOI are much smaller than the weights for EOL, but this is expected due to the differences in the definitions of these concepts as EOL is either present or not whereas the size of BOI varies and the weighting for BOI is for each unit of BOI. It is recommended that the weights generated using regressions without a QALY squared term ($MRS_{(1)}^{BOI}, MRS_{(1)}^{EOL}$) are preferable for use in policy, as these assume that the value of a QALY gain is equal regardless of the number of QALYs gained. The weights derived using regressions including a QALY squared term ($MRS_{(2)}^{BOI}, MRS_{(2)}^{EOL}$) allow that the societal value of a QALY gain changes as the number of QALYs gained changes. This means that these weights reflect a societal preference that values larger QALY gains proportionately lower, and hence are not only a reflection of the societal weighting for BOI or EOL. However for

both $MRS_{(1)}^{BOI}$ and $MRS_{(2)}^{BOI}$ the proportional weight given to each unit of BOI reduces as the size of QALY gain from treatment increases. For example for $MRS_{(1)}^{BOI}$ the weight for 10 units of BOI is 0.4 (10*0.040) regardless of whether QALY gain is 0.05, 0.5 or 5, but proportionately the weight for BOI differs relative to the size of the QALY gain. This means that BOI is proportionately less important when the QALY gain is large. This occurs as a result of the additive models used to model the data and is a potential limitation of this approach.

Limitations

An important concern with this survey is the use of an online sample and whether it is representative of the nation. An online sample 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. Respondents receive points for every survey they complete that can be exchanged for goods, which also may lead to the motivation for answering the survey to be questioned. The importance of these selection processes for the answers obtained in the survey is not known. However, the recruitment used 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 concept of BOI in the DCE task. This main survey was therefore designed to minimise this problem and comprised: an introduction video, practice questions involving a feedback screen 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, suggesting that respondents felt that they understood the questions that were asked. Robustness analyses excluding the first question for each individual had an impact on BOI coefficients in terms of significance or direction in only 2 of all 35 models, suggesting there were no significant learning effects in the study. However, excluding respondents who may not have understood or engaged with the survey did impact on the direction and significance of the coefficients for BOI (with only changes in magnitude for all other coefficients). Over one third (38.8%) of respondents chose to treat a patient group with a larger number of total lifetime QALYs after treatment, but smaller QALY gain from treatment and lower BOI before treatment than the other profile, suggesting that they may have misunderstood the DCE task. A further 10.1% of respondents did not see a question of this type. In robustness analyses the exclusion of the respondents

who were identified as possibly misunderstanding the task in this way had an impact on the results for BOI, affecting the significance and magnitude of the coefficients, but had little impact on the results for QALY gain and EOL. This has an impact on the results for BOI as these respondents were choosing to treat the patient group with lower BOI. It is possible that some of the respondents identified as possibly misunderstanding the task were expressing a genuine societal preference, yet this is unlikely to explain all of these responses. Overall the robustness analyses suggest that the coefficients and weights estimated for BOI using all data may be an underestimation as some respondents may have misunderstood the DCE task regarding BOI.

Overall, the results indicated that life expectancy without the condition in the profiles (N) impacted on the size of the coefficients. This 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)). This also has implications for the pooled models that combine the data across all life expectancies without the condition, and for the weights that are based on the pooled data.

In the attitude questions a large number of respondents chose to give equal priority to treating all patients, yet respondents were not allowed to give equal priority to treating both patient groups in the DCE task. This can be seen as a limitation of the DCE task as respondents had to choose whether the NHS should treat either patient group A or patient group B where the respondent may have been indifferent between treating either group. However forced choices for a small group of respondents should create random noise in the analysis rather than affect the results. In addition the attitudinal responses may also be reflecting a preference regarding equal access to health care when a patient is in need, rather than a rejection of BOI, EOL or TI. It must also be remembered that the attitudinal questions are dichotomous questions that do not involve trade-offs.

The experimental design of the DCE requires that the regression specifications are additive. All weights are therefore based on this assumption, yet the results may be affected by a possible interaction between the levels of BOI and QALY gains. The additive model also means that weights can be produced for interventions with positive BOI or EOL even when the QALY gain is zero. However it is unlikely in a policy context that weights would be required for an intervention providing zero QALY gains.

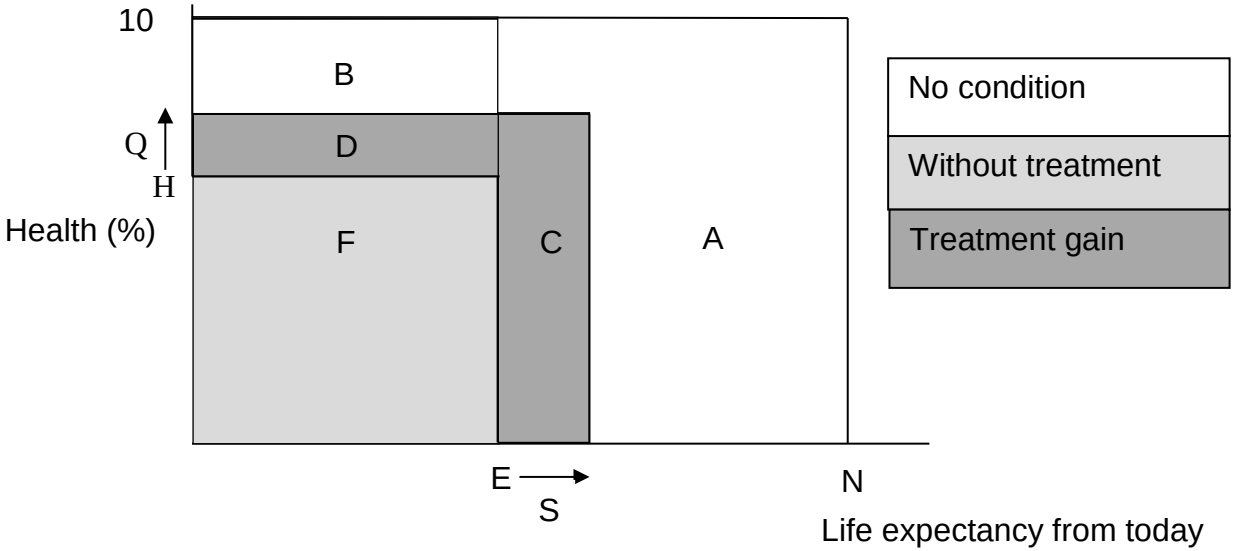
Other possible limitations are the assumption of zero time preference and the exclusion of age from the attributes and conceptual framework used in the survey. Furthermore the

concepts of EOL and TI are constrained by the definitions used in the survey. The NICE definition of end of life states under 2 years life expectancy whereas the models estimated here use 2 years or less due to the limited number of levels for the life expectancy variable. The NICE definition also requires that the condition is a 'rare' disease, but this has not been considered here.(15) Therapeutic improvement and innovation as outlined by DH may imply a treatment involving technological innovation, but the survey here has focussed solely upon the impact of the treatment on the size of QALY gain.(12)

CONCLUSION

This study provides the first attempt to operationalize the concept of BOI by combining the conventional notion of severity (in terms of poor health) with survival, using QALY loss attributable to the condition. It also provides evidence on societal preferences for EOL and TI. The results indicate general support for maximising QALY gains, with some support for BOI as a consideration when weighting QALYs. The evidence did not support the idea of TI. There is robust and consistent support for EOL in general, but this conceptually overlaps with BOI and the two should not be used together. Overall there seems to be a strong preference for larger QALY gain but at a diminishing rate. These results indicate that a QALY is not a QALY regardless of the burden of the disease or life expectancy and provides a basis for determining the appropriate weights.

Figure 1: Representation of profile used in survey



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

Figure 2 (pictures on next page): 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 **Without treatment** • will live for **10** years from today • with **50%** health **With treatment** • will live for **11** years from today • with **60%** health	**Patient group B** **Without treatment** • will live for **10** years from today • with **50%** 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 from today	

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 □

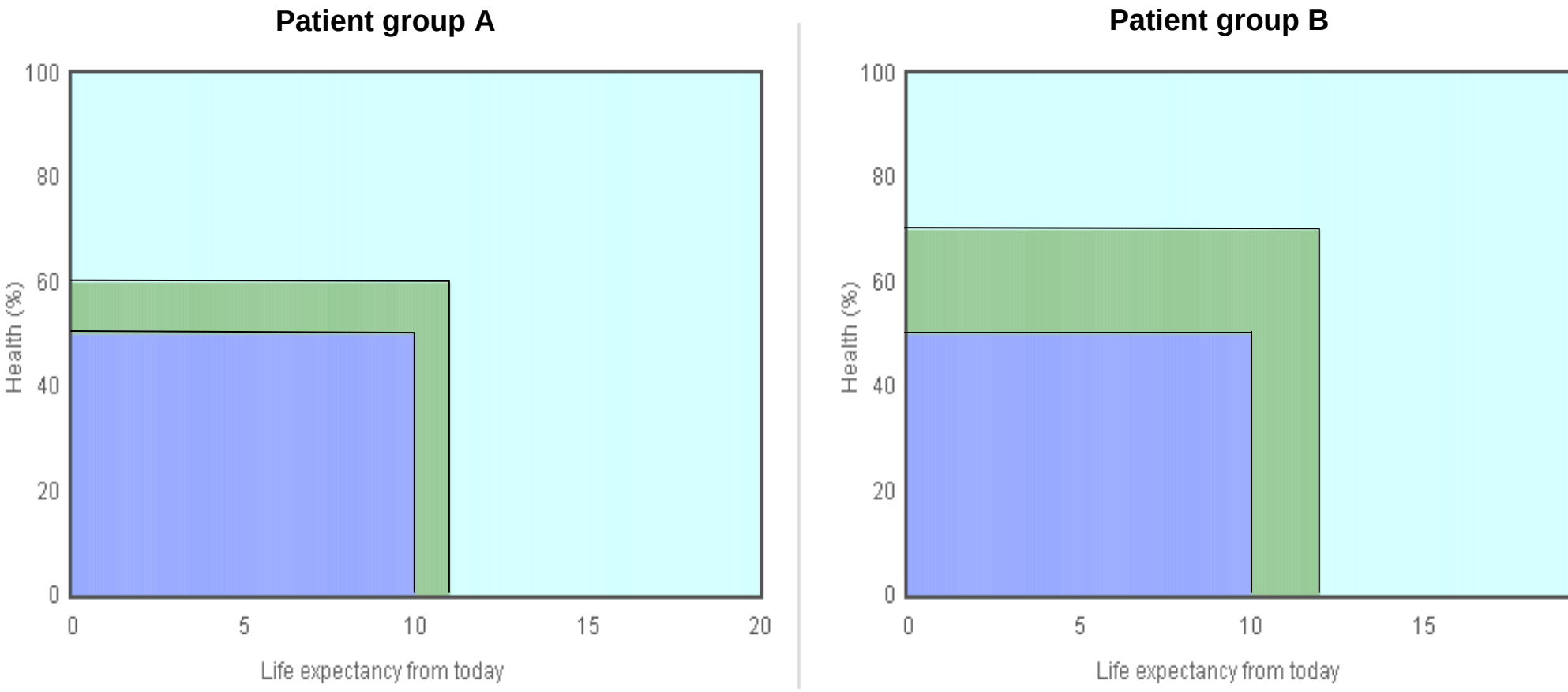

Patient group A
Patient group B
100
80
60
40
20
0
Health (%)
0
5
10
15
20
Life expectancy from today
Without treatment
Treatment gain
No condition

Figure 3: Practice Question 1 when normal life expectancy=20, feedback 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 **With treatment** • will live for **11** years from today • with **60%** health	**Patient group B** **Without treatment** • will live for **10** years from today • with **50%** 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 □

Table 1: Survey attributes and levels

Attribute	**Levels**	**Levels**	**Levels**	**Levels**
Life expectancy without the condition, 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 without 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
Number of pairs	160	120	140	160
Combinations of pairs (card blocs)	16	12	14	16

Table 2: Summary of statistics derived from the DCE design

	Normal life expectancy	5	20	40	80
QALY gain	Mean (s.d.)	0.673 (0.744)	2.039 (2.638)	3.012 (4.219)	8.510 (13.640)
	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)
	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)
	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)
	Minimum	1	12	16	32
	Maximum	4.975	19.975	39.975	79.975
Life expectancy without treatment	Mean (s.d.)	1.206 (1.229)	3.621 (3.585)	8.665 (10.696)	15.599 (20.582)
	Minimum	0.25	0.25	0.25	0.25
	Maximum	5	10	30	60

Table 3: Sociodemographic characteristics

	All respondents	England*
N	3669	
Mean age (s.d.)	46.5 (16.6)	NA
Age distribution		
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%	
Found DCE questions quite or very difficult to understand	7.6%	
Found attitudinal questions quite or very difficult to understand	6.6%	
Median completion time in minutes from consent to end of survey (Interquartile range)	21 (17-27)	

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). NA=Not available

Table 4: 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	1022	7.7%	92.3%	42	26.2%	73.8%	1022	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%
	All respondents	3669	8.3%	91.7%	179	22.3%	77.7%	3669	7.0%	93.0%
2	5	1022	41.9%	58.1%	77	68.8%	31.2%	1022	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%
	All respondents	3669	46.8%	53.2%	252	67.1%	32.9%	3669	50.8%	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. Respondents were allowed up to 7 attempts at each practice question before moving on automatically to the next question.

Table 5: Regression analysis

			Life expectancy without the condition			
	Variables	All variants	5yrs	20yrs	40yrs	80yrs
$V_{(1)}^{BOI}$	QALY	0.149***	1.813***	0.437***	0.191***	0.086***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	BOI	0.006***	0.068*	-0.015	0.028***	-0.003
		(0.005)	(0.057)	(0.328)	(0.000)	(0.156)
	Log likelihood	-22604	-5466	-4153	-5421	-5615
	Rho-squared	0.111	0.228	0.212	0.120	0.188
	AIC	45212	10936	8309	10847	11234
	BIC	45229	10950	8323	10861	11248
$V_{(2)}^{BOI}$	QALY	0.276***	3.641***	0.751***	0.404***	0.171***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	QALY_sq	-0.004***	-0.709***	-0.037***	-0.014***	-0.002***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	BOI	0.017***	0.120***	-0.000	0.039***	0.005*
		(0.000)	(0.001)	(0.999)	(0.000)	(0.068)
	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	43581	10348	8114	10519	10859
$V_{(3)}^{BOI}$	QALY	0.309***	3.626***	0.784***	0.434***	0.192***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	QALY_sq	-0.004***	-0.698***	-0.039***	-0.014***	-0.002***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	BOIQL	-0.027***	0.000	-0.071***	-0.012*	-0.020***
		(0.000)	(0.993)	(0.000)	(0.072)	(0.000)
	BOISU	0.009***	0.150***	-0.003	0.033***	-0.000
		(0.000)	(0.000)	(0.870)	(0.000)	(0.994)
	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	43021	10332	8062	10312	10729
$V_{(1)}^{EOL}$	QALY	0.156***	1.628***	0.455***	0.190***	0.088***

			Life expectancy without the condition			
	Variables	All variants	5yrs	20yrs	40yrs	80yrs
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	EOL	0.521***	0.871***	0.359***	0.479***	0.152***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
	Log likelihood	-22284	-5312	-4119	-5378	-5610
	Rho-squared	0.124	0.250	0.218	0.127	0.189
	AIC	44571	10627	8243	10761	11225
	BIC	44588	10642	8256	10775	11239
$V_{(2)}^{EOL}$	QALY	0.281***	3.230***	0.762***	0.400***	0.175***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	QALY_sq	-0.004***	-0.602***	-0.037***	-0.014***	-0.002***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	EOL	0.609***	0.607***	0.375***	0.576***	0.314***
		(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	42854	10233	8042	10433	10818
	Observations	73,380	20,440	15,200	17,780	19,960

Notes: P values in parentheses. * significant at 10%, ** significant at 5%; *** significant at 1%.
QALY – quality adjusted life year gains; BOI – burden of illness measured as QALY loss; BOIQL – QALY loss due to poor HRQOL; BOISU – QALY loss due to shorter life expectancy; EOL – life expectancy before treatment ≤2 years and survival gain≥ 3 months.

Table 6: Marginal rate of substitution for BOI and EOL by size of QALY gain ($MRS_{(2)}^{BOI}$, $MRS_{(2)}^{EOL}$)

QALY gain	0.05	0.1	0.5	1	2	5	10	20
$MRS_{(2)}^{BOI}$*	-0.063	-0.063	-0.063	-0.064	-0.066	-0.073	-0.087	-0.141
$MRS_{(2)}^{EOL}$**	-2.170	-2.173	-2.197	-2.229	-2.294	-2.516	-3.000	-4.875

*Change in QALY gains required to maintain the level of utility when 1 unit of BOI is lost.
** Change in QALY gains required to maintain the level of utility when moving from not being EOL to being EOL.

Table 7: 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%
	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%

	Normal life expectancy	5	20	40	80	All
	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 amount of benefit to a large number of patients	8.3%	11.3%	10.6%	8.3%	9.5%
	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					
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%
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%

Note: Respondents were instructed to choose which of the grouped statements they agreed with most.

Reference List

(1) Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, et al. Methods for the estimation of the NICE cost-effectiveness threshold. 2013.

(2) Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. Health Econ 2004 May;13(5):437-52.

(3) Weinstein MC. A QALY is a QALY--or is it? J Health Econ 1988 Sep;7(3):289-90.

(4) Nord E. Concerns for the worse off: fair innings versus severity. Soc Sci Med 2005 Jan;60(2):257-63.

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

(6) Green C. Investigating public preferences on 'severity of health' as a relevant condition for setting healthcare priorities. Soc Sci Med 2009 Jun;68(12):2247-55.

(7) Shah KK. Severity of illness and priority setting in healthcare: a review of the literature. Health Policy 2009 Dec;93(2-3):77-84.

(8) Dolan P, Edlin R, Tsuchiya A. The relative societal value of health gains to different beneficiaries. NICE Social QALY Project; 2008.

(9) Nord E. The trade-off between severity of illness and treatment effect in cost-value analysis of health care. Health Policy 1993 Aug;24(3):227-38.

(10) Hansson LF, Norheim OF, Ruyter KW. Equality, explicitness, severity, and rigidity: the Oregon plan evaluated from a Scandinavian perspective. J Med Philos 1994 Aug;19(4):343-66.

(11) Williams A. Intergenerational equity: an exploration of the 'fair innings' argument. Health Econ 1997 Mar;6(2):117-32.

(12) Department of Health. Value based pricing: impact assessment. Web archives National archives 2010Available from: URL: http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/en/consultations/liveconsultations/DH_122760

(13) Olsen JA. A note on eliciting distributive preferences for health. J Health Econ 2000 Jul;19(4):541-50.

(14) Rodriguez-Miguez E, Pinto-Prades JL. Measuring the social importance of concentration or dispersion of individual health benefits. Health Econ 2002 Jan;11(1):43-53.

(15) National Institute of Health and Care Excellence. Appraising life-extending, end of life treatments. London: NICE; 2009.

(16) Lancsar E, Wildman J, Donaldson C, Ryan M, Baker R. Deriving distributional weights for QALYs through discrete choice experiments. J Health Econ 2011 Mar;30(2):466-78.

(17) Brazier J, Rowen D, Mukuria C, Whyte S, Keetharuth A, Risa Hole A, et al. Eliciting societal preferences for burden of illness, therapeutic improvement and end of life for value based pricing. Universities of Sheffield and York; 2014. Report No.: EEPRU research report no 11.

(18) Carlsson F, Martinsson P. Design techniques for stated preference methods in health economics. Health Econ 2003 Apr;12(4):281-94.

(19) Kuhfield WF. Marketing Research Methods in SAS. SAS Institute Inc; 2005.

(20) Akaike H. Information, Theory and an Extension of the Maximum Likelihood Principle. In: Petrov BV, Csaki F, editors. Second International Symposium on Information Theory.Budapest: Akademai Kiado; 1973. p. 267-81.

(21) Schwarz G. Estimating the dimension of a model. Annals of Statistics 1978;6:461-4.

(22) Hole AR. A comparison of approaches to estimating confidence intervals for willingness to pay measures. Health Econ 2007 Aug;16(8):827-40.

(23) Dolan P. Modeling valuations for EuroQol health states. Med Care 1997 Nov;35(11):1095-108.

(24) Shah KK, Tsuchiya A, Hole AR, Wailoo A. Valuing health at the end of life: A stated preference discrete choice experiment. NICE DSU; 2012.