A022 Eliciting societal preferences for burden of illness, therapeutic improvement and end of life for value based pricing

Donna Rowen[a], John Brazier[a], Clara Mukuria[a], Sophie Whyte[a], Anju Keetharuth[a], Arne Risa Hole[b], Aki Tsuchiya[a,b], Phil Shackley[a]

[a] School of Health and Related Research (ScHARR) [b] Department of Economics, University of Sheffield

Aims: The Department of Health Value-based pricing (VBP) policy aims to assess the cost-effectiveness of medicines taking into account a broader scope of 'value', including the burden of disease and wider societal benefits. This paper reports on a study eliciting societal preferences for VBP across: (1) burden of illness (BOI) from a medical condition, defined as QALY loss due to premature mortality and morbidity; (2) therapeutic improvement (TI) defined as preferences for large QALY gains that are disproportionately larger than the size of gain; 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, health-related quality of life (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.

Data: 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.

Results: 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 and this was also found for models capturing proportional shortfall. 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.

Conclusions: The results support the argument that the social value of a QALY is not equal between recipients, but also depends on the burden of their illness and expected survival. However, there are concerns about the reliability of using an online sample.

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 typically assumes that a QALY is worth the same regardless of who gets it, yet there is emerging evidence that members of the public may value some QALY gains more highly than others depending on who receives them. The literature has uncovered a broad range of attributes across which the value of a QALY may be expected to vary including age, health state before treatment, size of the health benefit, socio-economic background, 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 basis can be surveys of the general public on the grounds that they are potential tax payers or on the basis of democracy. There is most support in the literature for the idea that those in worse health should be given greater priority than those in better health, often referred to as the 'severity argument' (Green, 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 (Nord, 1993; Nord, 2005), but severity of a condition is typically seen in terms of mortality as well as the quality of the health state. As argued by Hansson and colleagues (1994, p353) "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" (p353 ,Hansson et al 1994). 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 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).

In a consultation document, the Department of Health set out a new mechanism for pricing drugs in the UK known as Value Based Pricing (VBP) (DH, 2010). VBP 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 terms set out by Hansson and colleagues (1994) and

refers to this as burden of illness (BOI). BOI is measured as 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'. There is some evidence to suggest that there may be a threshold below which a given QALY gain has proportionately less value (Rodriguez-Miguez and Pinto Prades, 2002).

In addition there already exists another attribute that is used in health technology assessment 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 two years and a survival gain of 3 months or more (provided the condition is a 'rare' disease) (NICE, 2009). This attribute has been included in our 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 measured against life expectancy without the condition and/or reduced morbidity; 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 above. 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 (DCE), regression analyses of the survey and QALY weights.

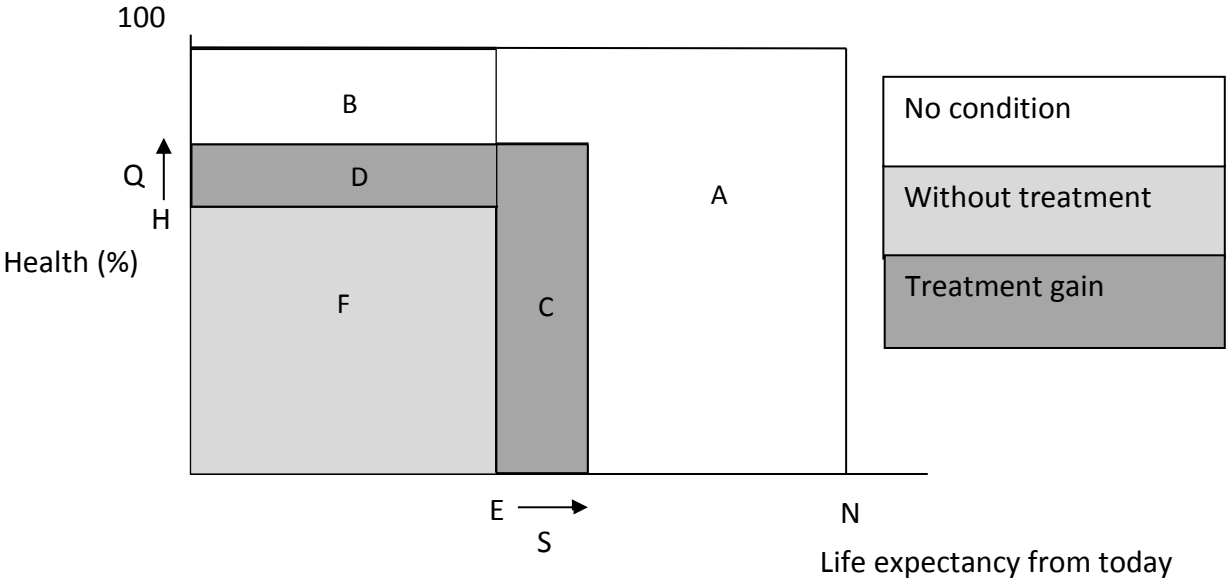
Methods

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 level of health without the disease is assumed to be 100% with life expectancy N. The expected 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 health loss from morbidity (areas B+D in Figure 1), and life expectancy loss from premature mortality (areas A+C in Figure 1), generated as 100*N - area F. QALY gain is areas D plus 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.

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

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 also already been successfully employed by Lancsar and colleagues (Lancsar et al, 2011). Several preparatory studies were undertaken to determine the choice of DCE and the online mode of administration: for further details see Brazier et al (2013).

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 are 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. 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 ensure precision over the more common characteristics of interventions in the UK where patients have a small number of years of life expectancy without the condition remaining and there are small QALY gains (Walker, 2011).

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 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

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 (Carlsson and Martinsson, 2003) and the true model specified in such a way as to allow for the estimation of an additive model including all parameters of interest. 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.

Analysis of data

The DCE data were modelled based on a random utility theory (RUT) framework. 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 any other alternative. Choices are based on a set of attributes captured in V_{ij} and other influencing factors that are not observed which are captured by the random component. DCE data provide the alternatives that individuals have chosen, in this case whether respondents thought the NHS should treat patient group A or patient group B, and these were modelled using the conditional logistic model which models the probability that individual *i* chooses profile *j* = A, B, for example, the probability of an individual choosing to treat patient group A (P_A) over B, given by:

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

V was modelled as a function of a vector of attributes *z* which 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 are 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 and these will also have a large BOI.

In addition respondents' choices may be driven by the relative sizes of the QALY gain or BOI, rather than their absolute value. Models are estimated to examine the notion of proportional shortfall described by Johannesson (2001) and Stolk et al. (2005), which represents the amount of QALYs lost due to a condition proportional to the patient's remaining QALY expectancy without the condition. This is estimated using:

- *stdBOI* representing proportional shortfall, generated using BOI divided by the QALY profile expected without the condition $\frac{BOI}{N}$
- *stdQALY* representing standardised QALY, generated using QALY gain divided by the QALY profile expected without the condition $\frac{QALY}{N}$

Model specification

The survey was designed to estimate an additive model, where the simple model for QALY gain and BOI is:

$$V_{(1)}^{BOI} = \beta_1 QALY + \beta_2 BOI + \varepsilon \quad (3)$$

Where V represents social value or utility and β represents the estimated coefficient. A more complex regression model including a QALY gain squared term to account for TI is:

$$V_{(2)}^{BOI} = \beta_1 QALY + \beta_2 BOI + \beta_3 QALY^2 + \varepsilon \quad (4)$$

Where a positive value for β_3 indicates an increasing marginal utility as QALY gains increase. The additive model specification was chosen to keep the model as simple and transparent as possible. Five model specifications are reported: (1) QALY gain and BOI, (2) QALY gain, QALY gain squared and BOI, (3) QALY gain, QALY gain squared and BOI split into BOI due to losses in health (BOIQL) and life expectancy (BOISU), (4) standardised QALY gain, QALY gain squared and BOI, (5) QALY gain, QALY gain squared and EOL.

Model performance

Performance of all regression models was assessed using the log-likelihood, Rho-squared, Akaike Information Criterion (AIC) and the Schwarz Bayesian Information Criterion (BIC). Models were preferred with larger log likelihood, larger Rho-squared and lower AIC and BIC. Collinearity was assessed using the variance inflation factor (VIF), with values greater than 10 indicating evidence of collinearity.

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 right sides of the screen). In addition respondents were excluded who 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 with patients who were most deserving of treatment. A final check was done by excluding the first survey question 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 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:

$$MRS^{BOI} = -MU_{BOI}/MU_{QALY} = -\frac{\partial U}{\partial BOI}/\frac{\partial U}{\partial QALY} \quad (5)$$

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 $V_{(1)}^{BOI}$ model specified in equation (3) this is:

$$MRS_{(1)}^{BOI} = -\frac{\widehat{\beta}_2}{\widehat{\beta}_1} \quad (6)$$

For the $V_{(2)}^{BOI}$ model specified in equation (4) this is:

$$MRS_{(2)}^{BOI} = -\frac{\widehat{\beta}_2}{\widehat{\beta}_1 + 2\widehat{\beta}_3 QALY} \quad (7)$$

MRS for EOL, MRS^{EOL}, is generated using equivalent regression specifications involving EOL rather than BOI. The regressions selected to generate the coefficients are estimated using all data collected in the survey, which uses 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 is to obtain a representation of the data across all levels of life expectancy without the condition.

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.

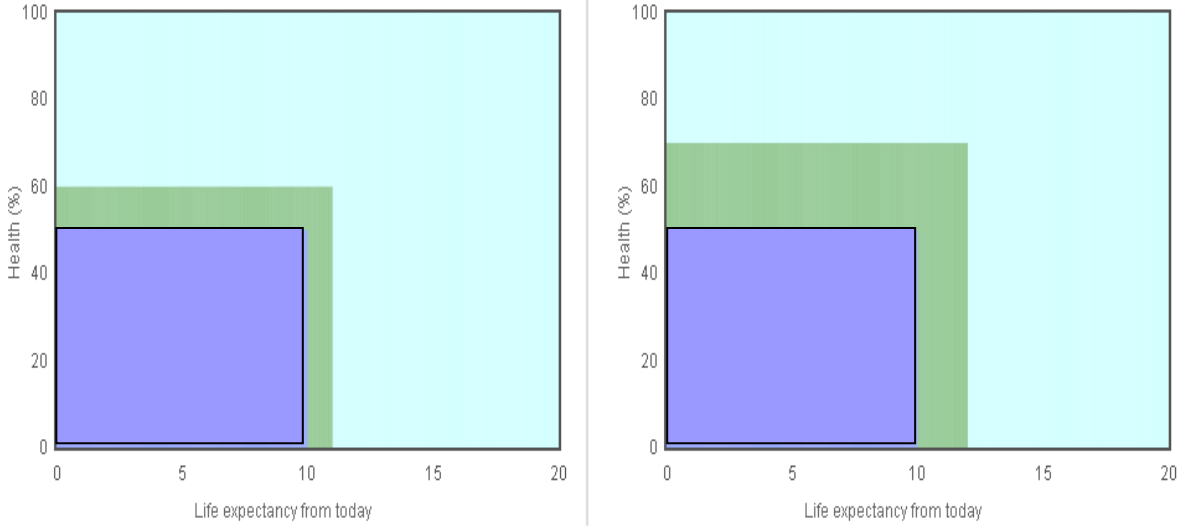
The survey had 2 practice questions which involved a “feedback screen” including an explanation of their choice with a chance for respondents to change their mind. The first practice question was dominant in QALY gain in one alternative while all other attributes were the same while the second was dominant in BOI in one alternative 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, and started the next

Figure 2: Practice Question 1 when life expectancy without the condition=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 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.

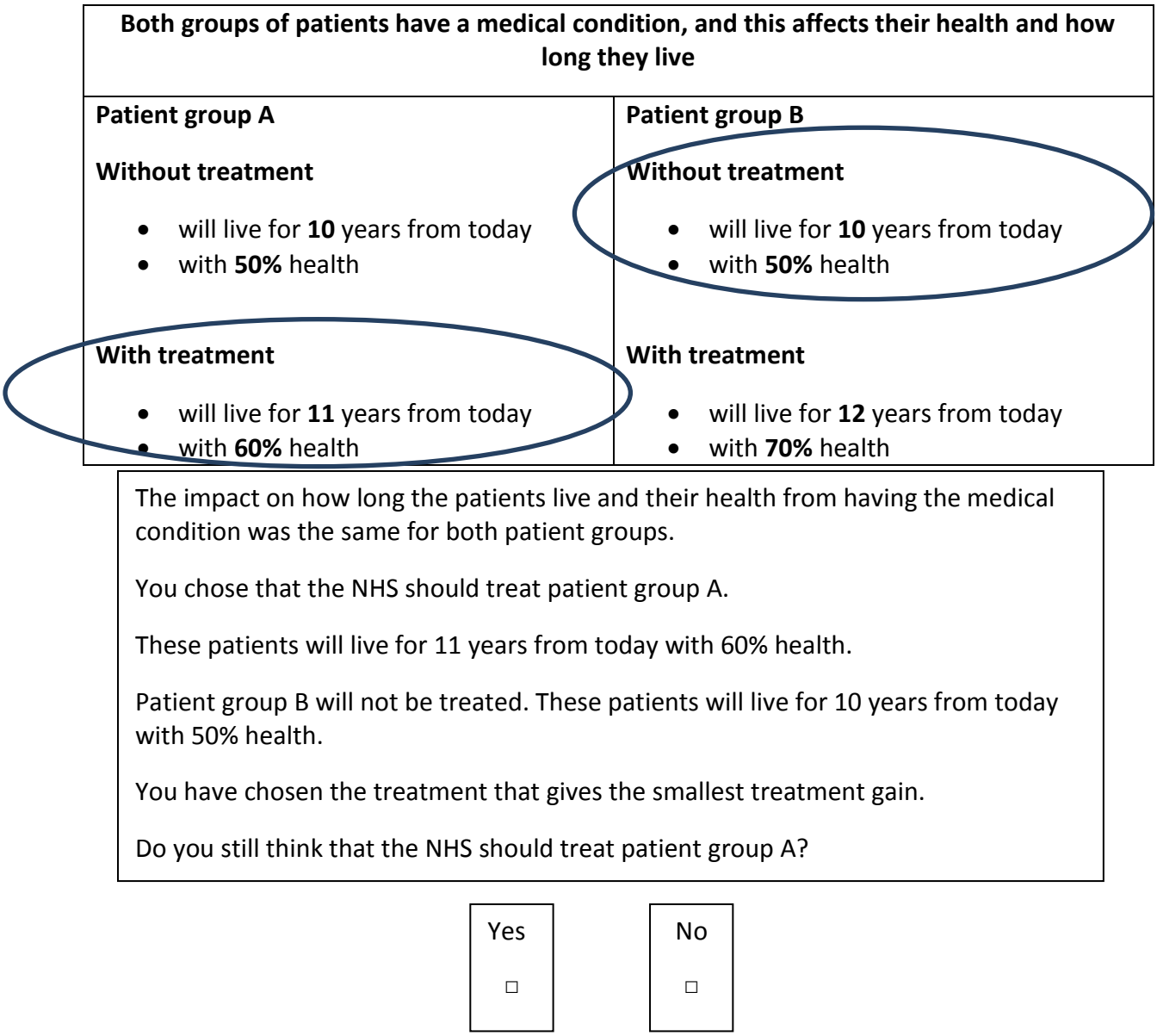
Remember that you can treat only 1 patient group.

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 3: Practice Question 1 when life expectancy without the condition=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 □

question if they did not change their mind, or were shown the question again from the first screen 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. Respondents were not offered an indifference or 'prefer not to say' option.

After the 2 practice questions respondents completed 10 DCE questions, 9 questions on attitudes, and 17 questions covering EQ-5D of own health, socio-demographics, understanding and what they thought of the survey. The ordering of the 10 DCE questions was random. 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 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 have been excluded from the main analysis. Characteristics of the sample were compared to the general population in England in Table 2. 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 than the general population of England, indicating poorer health. Although 66.9% of individuals stated their health in general was good or very good, and 37% stated that they were limited by a long term health condition or disability. A large proportion of the sample, 48.2%, had a degree or equivalent professional qualification.

Table 2: 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%
Degree or equivalent professional qualification	48.2%	NA
Health in general is very good or good	66.9%	NA
Limited by long term health condition or disability	37.0%	NA
EQ-5D score, mean (s.d.)	0.78 (0.26)	0.86 (0.23)†
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 (IQR)	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

RESULTS

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. 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.

Practice questions

In practice question 1, respondents overwhelmingly (93.0%) chose to treat the group with the highest QALY gain, other things the same, and this is consistent across the different variants, varying from 90.7% to 92.5%. In practice question 2 there is little evidence (50.8%) that, other things equal, respondents prefer 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. These final responses differed to their first response for 2-4% respondents, as respondents were offered the option to change their mind up to 7 times in each practice question. In robustness analyses excluding respondents identified as possibly misunderstanding the DCE task the proportion choosing to treat the patient group with higher BOI was 63.5% (remaining n=2247).

Regression results

Table 3 presents the regression results. The coefficients vary across the different variants meaning that the coefficients vary as life expectancy without the condition varies.

The coefficient for QALY gain was positive and significant across all models, indicating that respondents preferred profiles with higher QALY gains. The coefficient for the QALY gain squared term was negative and significant across all models, indicating that QALY gains were preferred at a decreasing rate and meaning that there was no support for TI: in fact the reverse was observed.

The coefficient for BOI was small, positive and significant, indicating that respondents preferred profiles with higher BOI, with the exception of the models estimated when life expectancy without the condition was 20 years. 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. 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 profiles with higher burden from life expectancy and lower burden in health. These coefficients are not consistent in direction between the questionnaire variants, though these coefficients are not significant where the sign is inconsistent. Standardising BOI does not reduce the

Table 3: Regression analysis using all data

	Variables	All	5yrs	20yrs	40yrs	80yrs
$V_{(1)}^{BOI}$	QALY	0.149***	1.813***	0.437***	0.191***	0.086***
	BOI	0.006***	0.068**	-0.015	0.028***	-0.003*
	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	45230	10951	8324	10862	11250
	Mean VIF	1.15	1.02	1.02	1.09	1.00
$V_{(2)}^{BOI}$	QALY	0.276***	3.641***	0.751***	0.404***	0.171***
	QALY_sq	-0.004***	-0.709***	-0.037***	-0.014***	-0.002***
	BOI	0.017***	0.120***	-0.000	0.039***	0.005**
	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
	QALY	0.309***	3.626***	0.784***	0.434***	0.192***
	QALY_sq	-0.004***	-0.698***	-0.039***	-0.014***	-0.002***
	BOIQL	-0.027***	0.000	-0.071***	-0.012**	-0.020***
	BOISU	0.009***	0.150***	-0.003	0.033***	-0.000
	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	14.994***	18.206***	15.018***	16.145***	13.689***
	stdQALY_sq	-14.803***	-17.716***	-14.905***	-22.436***	-13.118***
	stdBOI	0.758***	0.600***	-0.000	1.550***	0.360**
	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
$V_{(2)}^{EOL}$	QALY	0.281***	3.230***	0.762***	0.400***	0.175***
	QALY_sq	-0.004***	-0.602***	-0.037***	-0.014***	-0.002***
	EOL	0.609***	0.607***	0.375***	0.576***	0.314***
	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
	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; BOISU – QALY loss due to shorter life expectancy; BOIQL – QALY loss due to poor HRQOL; EOL – life expectancy before treatment ≤2 years and survival gain≥ 3 months; stdBOI - BOI divided by the QALY profile expected without the condition; stdQALY – QALY gain divided by the QALY profile expected without the condition.

differences between coefficients across variants and so the results do not seem to support the proportional shortfall model.

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.

Robustness

The robustness of results was examined as there were concerns that respondents who did not understand or engage with the survey may have 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 gain, BOI and EOL coefficients although there were variations in magnitude for all of them. The impact of excluding the first question for each individual had no impact on the coefficients in terms of significance and direction but there were small changes in magnitude.

Second, 1422 respondents who were identified as possibly having misunderstood the DCE task were excluded (see methods section for exclusion criteria). Some respondents did not answer a question that could be used to implement this exclusion criteria, and 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).

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. However the split variables were not robust following exclusions of respondents who arguably misunderstood the DCE task. There was no evidence of collinearity.

Marginal rate of substitution

The $MRS_{(1)}^{BOI}$ of 1 unit loss of BOI is -0.040 QALYs. $MRS_{(1)}^{BOI}$ indicates that if BOI increases by 1 unit, the level of utility is maintained if QALY gain decreases by 0.040 QALYs. This calculation assumes that the social value of a QALY 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 for a QALY gain of 1. However, the weighting for one extra unit of BOI now differs depending on the size of QALY gain as shown in Table 4, ranging from -0.063 to -0.141 as QALY gain changes from 0.05 to 20. The $MRS_{(1)}^{EOL}$ of EOL is -3.331 QALYs (regression coefficients not included here). $MRS_{(1)}^{EOL}$ indicates that by moving from not being EOL to being EOL, the level of utility is maintained by a QALY loss of 3.331. Allowing the value of a QALY to vary results in a range in $MRS_{(2)}^{EOL}$ of -0.2170 to -4.875 as QALY gain changes from 0.05 to 20.

Table 4: 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

Attitudinal questions

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 (see Table 5). The results suggest some support for BOI, though this depended on the size of treatment gain, and some support for EOL provided patients live at a reasonable level of health. 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. There was very little support for TI. However, it must be remembered that these are dichotomous questions and do not permit trading-off between attributes.

Table 5: Responses to the attitudinal questions

Q	**Response**	**%**
	BOI	
1	The NHS should give priority to treating patients who are very ill	40.7%
	The NHS should give the same priority to treating all patients who are ill, regardless of how ill they are	59.3%
2	The NHS should give priority to treating patients who are very ill and will die early because of their illness	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	57.5%
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	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	44.4%
	The NHS should give the same priority to treating all patients, regardless of how ill they are or when they will die	46.1%

8	The NHS should give priority to treating patients who are very ill and will die early because of their illness	10.6%
	The NHS should give priority to treating patients who will get the largest amount of benefit from treatment	44.6%
	The NHS should give the same priority to treating all patients	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	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	38.4%
	The NHS should give the same priority to treating all patients, regardless of how ill they are or when they will die	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	3.9%
	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	56.4%
	The NHS should give the same priority to treating all patients, regardless of how ill they are or when they will die	39.7%
7	The NHS should give priority to extending the life of patients expected to die soon	12.0%
	The NHS should give priority to treating patients who will get the largest amount of benefit from treatment	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	8.1%
	The NHS should give priority to treatments that give a small amount of benefit to a large number of patients	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	82.4%
	Combined	
9	The NHS should give priority to treating patients who are very ill and will die early because of their illness	13.1%
	The NHS should give priority to treating patients who will get the largest amount of benefit from treatment	52.1%
	The NHS should give priority to treating patients who will live for a long time and be in good health after treatment	34.8%

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

DISCUSSION

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. 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 end of life (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 more unit of BOI is equivalent to -0.04 QALYs, and EOL is equivalent to -3.331

QALYs (assuming the social value of a QALY does not change with size of QALY gain). 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 (DH, 2010). 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).

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, yet 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, while their larger follow-on study indicated little support for EOL (Shah et al, 2012), 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 survival 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 always have a large BOI. There is an important policy decision about whether to include one or the other, but not both. The advantage of BOI is that is 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. 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.

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 tasks set. 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. However, exclusion of respondents who were identified as possibly misunderstanding the task had an impact on the results for BOI, but had little impact on the results for QALY gain and EOL.

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). The results do not support the notion of a proportional shortfall approach to BOI proposed by Johannesson (2001) and Stolk et al. (2005). However, exclusion of respondents who were identified as possibly misunderstanding the task generated results that 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 life expectancy without the condition, meaning that in their experiment proportional shortfall could be capturing the proportional shortfall or the equivalent proportional gain, whereas in the DCE used here there is one variable capturing proportional shortfall and another capturing proportional gain.

The experimental design of the DCE requires that the regression specifications are additive. All weightings 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 analysis has not examined this. In addition, there are multiple variants of the questionnaire each with a different level of life expectancy without the condition (5, 20, 40 and 80 years) and multiple regression specifications, and the choice of variant and specification used to estimate the weights impacts on the results. The weightings reported here are based on the pooled sample.

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. In addition 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.

Conclusion

This study provides the first attempt to operationalize the concept of BOI by combining the conventional notion of severity (in terms of low health status) with survival, using QALY loss attributable to the condition. It also provides evidence on societal preferences for EOL and TI. The results indicate some support for BOI as a consideration when weighting QALYs. The evidence did not support the idea of TI. There was 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 can be used to estimate weights that can be applied to different QALY gains arising from new interventions and those that will be displaced in order to make way for them in a health service with a limited budget.

Acknowledgements: This study was commissioned by the UK Department of Health's Medicines, Pharmacy and Industry group, and undertaken by the Department of Health Policy Research Unit in Economic Evaluation of Health and Care Interventions (EEPRU). 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 and Jose Luis Pinto Prades for comments on earlier versions.

References

Brazier J, Rowen D, Mukuria C, Whyte S, Keetharuth A, Hole AR, Tsuchiya A, Shackley P. (2013) Eliciting societal preferences for burden of illness, therapeutic improvement and end of life for value based pricing: a report of the main survey. EEPRU Research Report 01/13

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 et al (2013) Methods for the estimation of the NICE cost effectiveness threshold. Centre for Health Economics Research Paper 81, University of York

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)

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

Hansson 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

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

Linley WG and Hughes DA (2013) Societal views on NICE, cancer drugs fund and value-based pricing criteria for prioritising medicines: A cross-sectional survey of 4118 adults in Great Britain. Health Economics 22: 948-964.

National Institute for Health and Clinical Excellence (NICE) (2009). Appraising life-extending, end of life treatments. 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. Soc Sci & Med 60: 257-263

Norman R, Hall J, Street D, Viney R. (2013) Efficiency and equity: a stated preference approach. Health Economics 22:568-581

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

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

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, E. A., Pickee, S. J., Ament, A. H., & Busschbach, J. J. (2005) Equity in health care prioritisation: an empirical inquiry into social value. Health Policy 74: 343-355

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