Eliciting Societal Preferences for Weighting QALYs for Burden of Illness and End of Life

Donna Rowen, PhD, MSc, John Brazier, PhD, MSc, Clara Mukuria, PhD, MSc, Anju Keetharuth, PhD, MBA, Arne Risa Hole, PhD, MSc, Aki Tsuchiya, PhD, MA, Sophie Whyte, PhD, MMath, Phil Shackley, PhD, MSc

Objectives. *Recent proposals for value-based assessment, made by the National Institute of Health and Care Excellence (NICE) in the United Kingdom, recommended that burden of illness (BOI) should replace end of life (EOL) as a factor for consideration when deciding on new health technologies. This article reports on a study eliciting societal preferences for 1) BOI from a medical condition, defined as quality-adjusted life year (QALY) loss due to premature mortality and prospective morbidity, and 2) EOL, defined as expected life expectancy of less than 2 years and expected life expectancy gain from new treatment of 3 months or more.* ***Methods***. *A discrete choice experiment survey was conducted with an online UK general population sample. Respondents chose whether they thought the health service should treat patient group A or B: life expectancy and health-related quality of life (HRQOL) with current treatment or life expectancy and HRQOL gains from new treatment, respectively. These attributes were used to derive BOI, QALY gain, and EOL. The respondents' choices were analyzed using conditional logistic regression with a range of specifications examined, including BOI or EOL, QALY gain and QALY gain squared, and robustness. QALY weights were estimated.* ***Results***. *The sample of 3669 respondents was representative of the UK population for age and sex. QALY gain had a positive and significant coefficient across all models. QALY gain squared term was negative and significant across all models, indicating a diminishing marginal social value from QALY gains. When included, the BOI coefficient was generally small, positive, and significant, but this was not consistent across the different life expectancy variants. EOL was always positive and significant.* ***Conclusions***. *The social value of a QALY gain is not equal between recipients but depends on whether they are end of life, and it may depend on the prospective burden of illness.* ***Key words:*** *QALYs; burden of illness; severity; end of life.* ***(Med Decis Making 2016;36:210–222)***

Received 18 December 2014 from School of Health and Related Research (ScHARR) (DR, JB, CM, AK, AT, SW, PS) and Department of Economics (ARH, AT), University of Sheffield, Sheffield, United Kingdom. Revision accepted for publication 8 July 2015. We are grateful to the survey respondents and to Gavin Roberts and Danny Palnoch from DH for their comments throughout this project. We also thank Angela Robinson for her input into the survey framework, as well as Emily Lancsar, Rachel Baker, Cam Donaldson, Jose Luis Pinto Prades, and Richard Cookson for comments on earlier versions. An earlier version of the manuscript was presented at the Health Economics Study Group in Sheffield, January 2014. Only the authors are responsible for any remaining errors. Financial support for this study was provided by the Policy Research Programme in the Department of Health. The study was undertaken by the Policy Research Unit in Economic Evaluation of Health and Care interventions (EEPRU) funded by the Department of Health Policy Research Programme. The views expressed are not necessarily those of the department. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. Revision accepted for publication 3 August 2015.

DOI: 10.1177/0272989X15619389

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 measure the cost-effectiveness of an intervention, in terms of the incremental cost per quality-adjusted life year (QALY), and compare this to threshold cost per QALY to reflect displaced activities.[1] This approach is designed to improve the efficiency of health care spending and 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, such as the age of the patient or the severity of the disease. These considerations have been made explicitly or have been shown to do so implicitly in their decision making.[2]

Address correspondence to John Brazier, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK; telephone: +44114 222 0726; fax: +44114 272 4095; email: j.e.brazier@sheffield.ac.uk.

QALY gains can be allocated different weights depending on the characteristics of the patient group (such as their age or severity of condition), meaning that some patients receive a higher weight while others receive a lower weight.[3] There is evidence to suggest that a QALY gain to younger patients or those in a poor state of health may be weighted more highly than a QALY gain to older patients or those in a better state of health, respectively.[2] This means that interventions that are more expensive can be found to be cost-effective if they treat a patient group whose characteristics entail that their QALY gains receive a higher weight.

The literature has uncovered a broad range of attributes across which the value of QALY gains may be expected to vary, including age, health state with current treatment, socioeconomic background of a typical patient, degree of responsibility, and broader notions of fair innings. Arguably, 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 and/or the basis of democratic principles. There is emerging evidence that members of the public weigh some QALY gains more highly than others depending on who receives them.[4–8] A key finding is that those in worse health should be given greater priority than those in better health, often referred to as the ''severity argument,'' although this is not found in all studies.[8]

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 (HRQOL) 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 of the same size at the higher end of the scale. However, there is no reason to limit the severity of a condition to the impact on HRQOL. As argued by Hansson and colleagues, ''If health benefit with treatment is measured along the axes of mortality, pain, physical, mental and social functions, so should severity of disease.''[10(p353)] In other words, severity can be measured in terms of HRQOL and life expectancy.

There are two perspectives that can be taken in relation to severity. One is to take into account the health profile of a person's entire lifetime, and the second is to limit the perspective to prospective health. Using prospective health means not directly taking into account previous health experience. There is evidence to suggest that when people are asked to make choices over hypothetical scenarios, they choose on the basis of overall lifetime health prospects rather than future health prospects alone.[5,11] Conversely, 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 article attempts to operationalize the notion of ''burden of illness (BOI),'' which is the severity of a disease measured in terms of its impact on HRQOL and life expectancy, based on prospective and not lifetime health. BOI was originally set out by Hansson and colleagues.[10] The argument for using BOI is that it is compatible with cost per QALY analysis (see also Department of Health[12]) because it uses prospective QALY loss suffered by patients relative to their prospects in the absence of the disease from the decision point. Furthermore, the QALY loss is calculated according to current treatment. It is important to note that this form of BOI differs to some other uses in public health, as it is not concerned with the prevalence of the disease.

There is currently no evidence on societal preferences for this notion of BOI. The motivation for this study was to elicit evidence from the general public on their preferences for BOI. This research is policy relevant in the United Kingdom, as National Institute of Health and Care Excellence (NICE) proposed to take BOI into consideration in its appraisals of health technology assessments under the new policy of value-based assessment, to replace the current inclusion of the ''end-of-life'' (EOL) criterion.[13] In addition, BOI is a consideration in health technology assessment for the National Health Care Institute in the Netherlands (see van de Wetering et al.[14] for an overview).

This study also examines the 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 life expectancy gain from new treatment of 3 months or more (provided the condition is a ''rare'' disease).[15] Recent evidence suggests that societal preferences do not support EOL,[11,16,17] with the exception of one small UK study that found weak support for EOL.[18] This study provides further evidence on the preferences of members of the general public for EOL.

In addition to concerns for BOI and EOL, there is also a possibility that the marginal social value of QALYs gained diminishes as the size of QALY gain increases. For a given total of QALYs gained across people, not only the size of this total may affect social

welfare but also how it is distributed across people. It may be more preferable to disperse a given amount of benefits thinly across more patients rather than concentrate the benefit in the form of larger gains for fewer patients. This is the third criterion examined in this study.

In summary, the aim of this research was to elicit societal preferences (as opposed to personal preferences) of members of the public for 1) BOI as defined above, 2) EOL as defined by NICE, and 3) dispersing or concentrating QALY gains across the population. This article 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 of this research are intended to inform the application of incremental cost per QALY analyses and not individual patient-level decisions.

METHODS

The Framework

Members of the UK general population were surveyed online and asked to indicate which of 2 hypothetical patient groups they thought the National Health Service (NHS) should treat. The 2 patient groups were described in terms of life expectancy with current treatment, HRQOL with current treatment, life expectancy gain from new treatment, and HRQOL gain from new treatment. These 4 components were presented to respondents on a diagram like Figure 1, where a prospective health profile with current treatment is represented by HRQOL or health status (h), and life expectancy (e) is measured from the point at which the decision about new treatment is being considered. To estimate BOI in terms of the QALY loss associated with the condition, it is necessary to establish an expected or target level of HRQOL and life expectancy. In this study, the expected profile without the condition is assumed to be 100% HRQOL with life expectancy, n; 100% was chosen to make it simpler for respondents rather than as a realistic figure. The improvement from the new treatment is represented by a gain in HRQOL, q, and an improvement in life expectancy, s. BOI is the loss of HRQOL and life expectancy from their expected or target levels with current treatment, 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

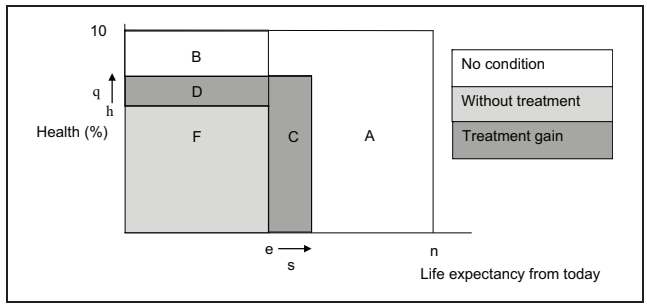

Figure 1 Representation of profile used in survey. n = life expectancy without the condition, e = life expectancy without treatment, s = life expectancy gain from treatment, h = health-related quality of life (HRQOL) before treatment, and q = HRQOL gain from treatment.

gain is areas D+C, and end of life is where e is 2 years or less and s is 3 months or more. BOI and EOL are similar in that they take into account a person's prospective health profile. EOL is more limited because it only considers length of life and operates as a dichotomous variable; either a patient group meets the criterion or it does not. BOI is a continuous variable and also takes in account HRQOL. However, they are correlated as a patient who meets EOL with less than 2 years to live is more likely to have a high burden of disease. Nevertheless, the correlation is far from perfect; for example, a very elderly group of patients may not have many years to live even without the medical condition, and so the BOI is quite low but they are at the EOL.

Elicitation Technique

A discrete choice experiment (DCE) based on pairwise comparisons was chosen to elicit preferences because it permits the simultaneous consideration of different attributes in a format that is amenable to online administration. 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 tradeoff questions, a qualitative survey to further explain the findings of the large preparatory online survey, and a 6-arm online and face-to-face survey examining different framings of questions and mode of administration.[19,20]

Selection of Attributes and Levels

The pairwise comparison DCE had 4 prospective attributes: life expectancy with current treatment

Table 1 Survey Attributes and Levels

Variant: Life Expectancy without the Condition, N	5 years	20 years	40 years	80 years
Attribute	Levels	Levels	Levels	Levels
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
Life expectancy 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
HRQOL without treatment (%), H	10, 20, 40, 60, 80	10, 20, 40, 60, 80	10, 20, 40, 60, 80	10, 20, 40, 60, 80
HRQOL 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
Design				
Number of pairs	160	120	140	160
Combinations of pairs (card blocs)	16	12	14	16

(e), life expectancy gain from new treatment (s), HRQOL with current treatment (h), and HRQOL gain from new treatment (q) (Figure 1). There were 4 different DCE designs (Table 1), each with a different level of life expectancy without the condition (*n*): 5 years, 20 years, 40 years, or 80 years. As a result, there were 4 variants of the questionnaire. Respondents saw one questionnaire variant (i.e., one level of life expectancy without the condition). Attribute levels and n were selected to cover a full range of potential levels. To ensure precision over the more common characteristics of interventions evaluated using health technology assessment, some alternatives had a small number of years of life expectancy remaining without the condition, with small QALY gains.

DCE design

A full factorial design would result in too many profiles of attribute combinations to be amenable to valuation. Profiles were selected using a D-optimality algorithm[21,22] 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). Each experiment was designed to minimize the amount of correlation between the derived variables. The designs also determined which of the paired profiles were displayed on the left or right of the computer screen. Impossible profiles, such as profiles involving HRQOL after new treatment of more than 100%, were excluded from the candidate sets for the designs (see Table 1 for design and Brazier et al.[23] for further details). In total, the DCE designs constituted 580 pairs of profiles, with the number of pairs varying across designs depending on the number of attributes and levels in the design. Pairs were allocated into 58 combinations (also known as ''blocks'') of 10 pairs. Summary statistics of the profiles generated by the DCE design are reported in Table 2 for each level of life expectancy. These show the large range of QALY gains considered in this survey, starting from 0.005 up to 63, and BOI from 1 to 80 QALYs lost and life expectancy from 0.25 to 60 years.

Analysis of Data

The DCE data were modeled based on a random utility theory (RUT) framework.[24,25] Within the RUT framework, ''utility'' U_{ij} for an individual i or, in the current context, the individual's judgement of societal value, is assumed to be a function of an explainable utility component V_{ij} and a random component ε_{ij}:

Table 2 Sociodemographic Characteristics

Characteristic	All respondents	England[a]
N	3669	
Age, mean (SD), y	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 (SD)	0.78 (0.26)	0.86 (0.23)[b]
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	
Completion time in minutes from consent to end of survey, median (interquartile range)	21 (17–27)	

Note: EQ-5D is scored using preference weights from reference 26. DCE, discrete choice experiment; NA, not available.
a. Statistics for England in the Census 2001. Questions used in this study and the census are not identical. The census includes persons aged 16 years and older, whereas this study only surveys persons aged 18 years and older. Age distribution is here reported as the percentage of all adults aged 18 years and older.
b. Interviews conducted in the Measurement and Valuation of Health (MVH) study.[32]

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

where j represents the alternatives individuals had within a choice set. The alternative chosen by the individual is assumed to confer greater utility, or in this case social value, than the other alternative. Choices are based on a set of attributes captured in V_{ij}, and other unobserved influencing factors are 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 modeled using the conditional logistic model, which models the probability that individual i chose profile $j = A, B$; for example, 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 social value from choosing to treat patient groups 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 modeled as a function of attributes z:

$$V = f(z), \quad (3)$$

where z is made up of a continuous variable *BOI*, representing burden of illness from both premature death (BOISU, A+C in Figure 1) and HRQOL loss (BOIQL, B+D in Figure 1) generated using $n - \frac{h}{100}e$, a continuous variable representing QALY gain from life expectancy (C in Figure 1) and from improved HRQOL (D in Figure 1) generated using $s\left(\frac{h+q}{100}\right) + \frac{q}{100}e$ and a dummy variable *EOL* representing end of life, using the NICE definition of 2 years

life expectancy or less ($e \leq 2$ years) and life expectancy gain from new treatment of 3 months or more ($s \geq 3$ months). It should be noted that QALY gain (C+D) is a part of BOI, although there is an independent component of BOI (A+B) and so they could be entered into the model together. However, BOI and EOL were not included in the same model specification due to conceptual overlap in these variables: EOL profiles will also have a large BOI.

Model Specification

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

$$V = \beta_{ij}\mathbf{z}_{ij} + \gamma_{ij}\mathbf{w}_{ij} + \varepsilon, \quad (4)$$

where V represents social value, $\mathbf{z}$ represents a vector containing the variables described above, and $\mathbf{w}$ represents the squared terms of QALY gain and BOI. This additive model specification was chosen to keep the model as simple and transparent as possible.

The first model for QALY gain and BOI is

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

A more complex regression model including a QALY gain squared term to account for nonlinear preferences is

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

where a negative value for β_3 indicates a diminishing marginal social value as QALY gains increase. The squared QALY term provides some evidence for or against the notion that respondents favor dispersing or concentrating the benefits. A squared term for BOI does not have a similar rationale and is not presented here, although it has been examined.

Different model specifications are reported representing models examining BOI and QALY gain (model $V_{(1)}^{BOI}$); BOI, QALY gain, and QALY gain squared (model $V_{(2)}^{BOI}$); EOL and QALY gain (model $V_{(1)}^{EOL}$); and EOL, QALY gain, and QALY gain squared (model $V_{(2)}^{EOL}$). All models were estimated using the clogit command with robust standard errors in Stata version 11.[27]

Model Performance

Performance was assessed using the log-likelihood, ρ^2, Akaike information criterion (AIC), and the Schwarz Bayesian information criterion (BIC).[28,29] Models were preferred with higher log-likelihood, larger ρ^2, 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: those who reported they found the survey quite or very difficult, those with a completion time of <5 minutes or >60 minutes, those who selected to treat the same patient group for all questions (respondents may be selecting either all left or all right sides of the screen), and exclusion of the first and last question (these responses may be less reliable if there were learning or fatigue effects).

Estimating Weights

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

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

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

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

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

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

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 across all variants of the questionnaire (5, 20, 40, and 80 years of life expectancy without the condition) to obtain a representation of the entire data set and to produce one set of consistent weights. The standard error (SE)

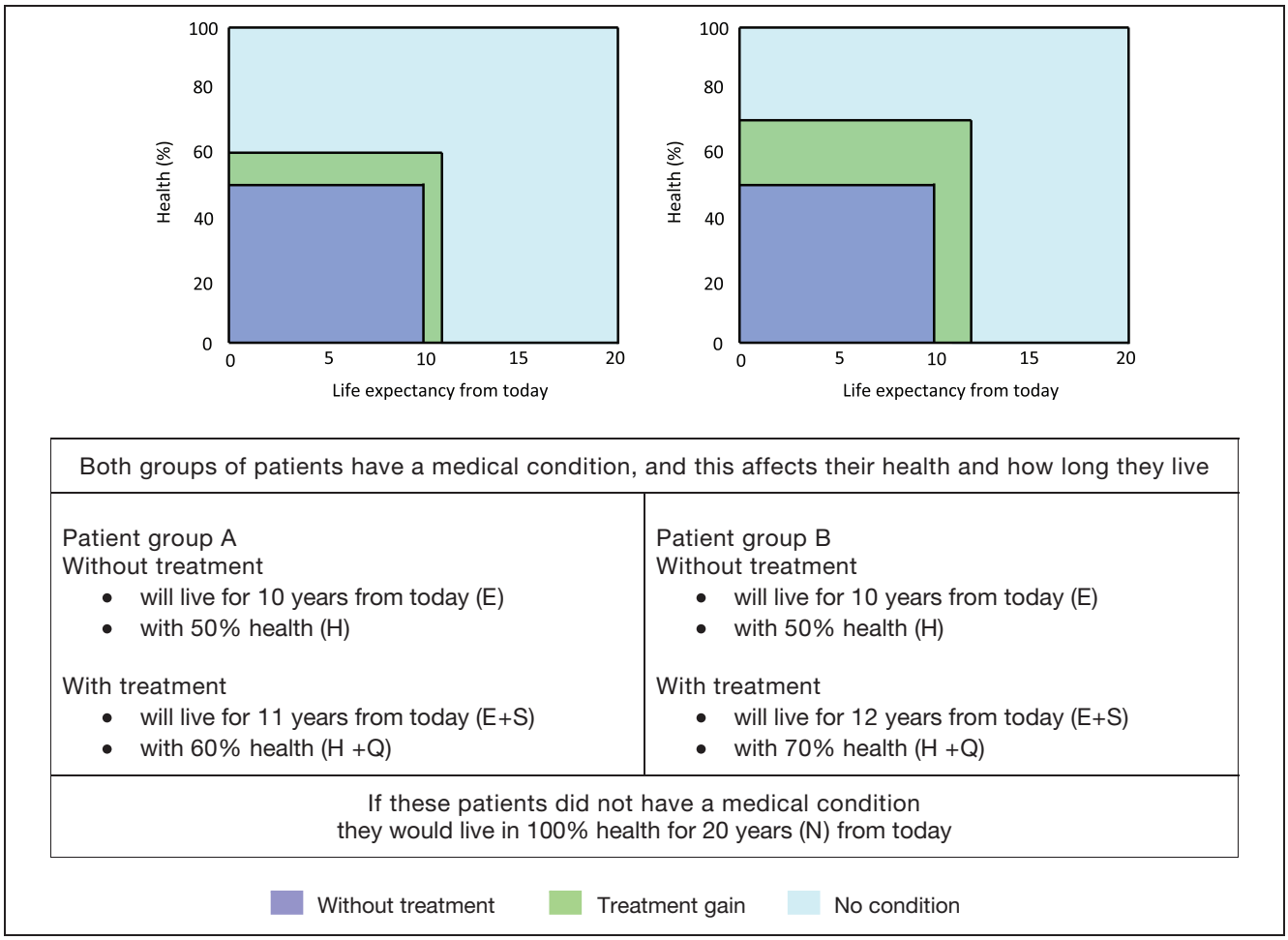

Figure 2 Discrete choice experiment (DCE) example question.

and the 95% confidence interval (CI) of the MRS were calculated using the Delta method.[30]

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 sex. Respondents read an information page, gave informed consent, were shown a short video explaining the questions, and completed 2 practice questions that involved a ''feedback screen,'' including an explanation of their choice with a chance for respondents to change their mind. 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 first practice question had one alternative with a larger QALY gain than the other alternative, ceteris paribus. The second practice question had one alternative with a larger BOI, ceteris paribus.

After the 2 practice questions, respondents completed 10 DCE questions (see Figure 2 for an example), 9 questions on attitudes, and 17 questions covering EQ-5D of their own health, sociodemographics, difficulty of understanding the questions, and what they thought of the survey. Respondents were randomly allocated to different questionnaire variants and combinations of profiles within each variant. The ordering of the 10 DCE questions was random for each respondent. Attitude questions were included to determine respondents' general views on BOI and EOL, and constant marginal value of gains from new treatment are reported elsewhere.[23]

RESULTS

The Data

A total of 3669 respondents completed the online survey, providing a completion 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 (Table 2). In comparison to the general population of England, the sample was largely representative for age and sex 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,[31] indicating poorer health. Although 66.9% of individuals stated their health in general was good or very good, 37% stated they were limited by a long-term health condition or disability and 33.6% stated they had experienced a serious illness themselves. A large proportion of the sample, 48.2%, had a degree or equivalent professional qualification. The majority of respondents, 77.7%, reported that the DCE questions were either very or fairly easy to understand.

Practice Questions

In practice question 1, which differed only in terms of QALY gain, respondents overwhelmingly (93.0%) chose to treat the group with the highest QALY gain. In practice question 2, which differed only in BOI, there was little evidence (50.8%) that respondents preferred to treat the patient group with higher BOI.

Modeling DCE Data

Across all models, QALY gain had a positive and significant coefficient, indicating that respondents preferred profiles with higher QALY gains (Table 3). The coefficient for the QALY gain squared term, when included, was negative and significant across all variants and models, indicating diminishing marginal social value of QALY gains.

The BOI coefficient, when included, was generally small, positive, and significant. However, this was not consistent across the different life expectancy variants, where BOI was nonsignificant for the 20-year variant in models $V_{(1)}^{BOI}$ and $V_{(2)}^{BOI}$, as well as for the 80-year variant in model $V_{(1)}^{BOI}$. BOI squared was tested but did not improve the models; 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. The coefficient for EOL was positive and significant, indicating support for EOL. This meant that respondents gave greater weight to shorter life expectancy with current treatment when life expectancy gains from new treatment were greater than 3 months.

The coefficients of all variables differed across the 4 different levels of life expectancy without the condition (see Table 3).

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 ρ^2 were the specifications with EOL.

Robustness of Results

The consequences of excluding the following were examined: 279 individuals who reported they had difficulty understanding the DCE questions, 208 individuals with a completion time of <5 minutes or >60 minutes, 23 individuals who chose the same option for all DCE questions, responses to the first DCE nonpractice questions, and responses to the last DCE nonpractice questions. The exclusions affected the magnitude of all coefficients but only affected their significance for a small number of models containing BOI, where the impact was not systematic.

Weights

The MRS was calculated by pooling across the 4 levels of life expectancy without the condition. The $MRS_{(1)}^{BOI}$ of 1 unit of BOI is –0.040 QALYs gained (95% CI: –0.068, –0.013). This indicates that if BOI increases by 1 unit, the level of social value is maintained by a reduction in QALY gain of 0.040. In this context, we are not describing individual utility in the usual sense but as a social value.

This calculation assumes that the social value of a QALY gain does not change with size of QALY gain. A significant QALY squared term suggests this assumption does not hold, and the MRS for 1 extra unit of BOI now differs depending on the size of QALY gain (Table 4), ranging from –0.063 to –0.141 as QALY gain changes from 0.05 to 20.

The $MRS_{(1)}^{EOL}$ of shifting from not being EOL to being EOL is –3.331 QALYs (95% CI: –3.711, –2.950). This indicates that by moving from not being

Table 3 Regression Analysis

			Life Expectancy without the Condition			
	Variables	**All Variants**	**5 Years**	**20 Years**	**40 Years**	**80 Years**
$V^{BOI}_{(1)}$	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	–22,604	–5466	–4153	–5421	–5615
	ρ^2	0.111	0.228	0.212	0.120	0.188
	AIC	45,212	10,936	8309	10,847	11,234
	BIC	45,230	10,951	8324	10,862	11,250
$V^{BOI}_{(2)}$	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	–21,775	–5160	–4043	–5246	–5416
	ρ^2	0.144	0.272	0.232	0.149	0.217
	AIC	43,555	10,326	8093	10,498	10,838
	BIC	43,582	10,350	8116	10,521	10,861
$V^{BOI}_{(3)}$	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	–21,489	–5148	–4013	–5138	–5346
	ρ^2	0.155	0.273	0.238	0.166	0.227
	AIC	42,987	10,303	8034	10,284	10,700
	BIC	43,024	10,335	8064	10,315	10,731
$V^{EOL}_{(1)}$	QALY	0.156***	1.628***	0.455***	0.190***	0.088***
		(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	–22,284	–5312	–4119	–5378	–5610
	ρ^2	0.124	0.250	0.218	0.127	0.189
	AIC	44,571	10,627	8243	10,761	11,225
	BIC	44,590	10,643	8258	10,776	11,241
$V^{EOL}_{(2)}$	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	–21,411	–5103	–4008	–5203	–5395
	ρ^2	0.158	0.280	0.239	0.156	0.220
	AIC	42,829	10,213	8022	10,411	10,797
	BIC	42,857	10,236	8045	10,435	10,820
	Observations	73,380	20,440	15,200	17,780	19,960

Note: *P* values in parentheses. AIC, Akaike information criterion; BIC, Schwarz Bayesian information criterion; BOI, burden of illness measured as QALY loss; BOIQL, QALY loss due to poor health-related QOL; BOISU, QALY loss due to shorter life expectancy; EOL, life expectancy before treatment ≤2 years and life expectancy gain from treatment ≥3 months; QALY, quality-adjusted life year gains.
*Significant at 10%. **Significant at 5%. ***Significant at 1%.

Table 4 Marginal Rate of Substitution for BOI and EOL by Size of QALY Gain ($MRS^{BOI}_{(2)}$, $MRS^{EOL}_{(2)}$)

	QALY Gain							
	0.05	**0.1**	**0.5**	**1**	**2**	**5**	**10**	**20**
$MRS^{BOI}_{(2)}$ [a]	–0.063	–0.063	–0.063	–0.064	–0.066	–0.073	–0.087	–0.141
$MRS^{EOL}_{(2)}$ [b]	–2.170	–2.173	–2.197	–2.229	–2.294	–2.516	–3.000	–4.875

Note: BOI, burden of illness measured as QALY loss; EOL, life expectancy before treatment ≤2 years and life expectancy gain from treatment ≥3 months; MRS, marginal rate of substitution; QALY, quality-adjusted life year gains.
a. Change in QALY gains required to maintain the level of utility when 1 unit of BOI is lost.
b. Change in QALY gains required to maintain the level of utility when moving from not being EOL to being EOL.

EOL to being EOL, the social value is maintained by a reduced QALY gain of 3.331. Including the QALY squared term allowed results ranging in $MRS^{EOL}_{(2)}$ of –2.170 to –4.875 as QALY gain changes from 0.05 to 20.

DISCUSSION

This was the first study to examine societal preferences for BOI alongside EOL. It was a large DCE survey using an existing online panel from the UK general population. Respondents preferred to treat patients with larger QALY gains, but at a diminishing rate. They also preferred to treat patients at the EOL. The results for BOI, defined as QALY loss due to premature mortality and prospective morbidity, were less robust across variants of the questionnaire but suggested some modest support for BOI. Using the MRS to estimate weights based on the pooled data set indicated that 1 unit of BOI is equivalent to 0.040 QALYs gained, and EOL is equivalent to 3.331 QALYs gained.

The diminishing marginal value of QALY gains indicates a preference to disperse QALY gains. Although not directly comparable in terms of the attributes included, recent UK and Australian studies also found QALY gains to have a positive and statistically significant impact, but at a declining rate.[32,33] Although the evidence on whether respondents prefer to concentrate or disperse gains is somewhat mixed, respondents do often favor greater dispersion.[34] The results are also consistent with studies finding evidence for dispersing life year gains above a certain threshold,[35,36] although our survey was not designed to examine this.

The evidence of a preference for EOL is consistent with a small UK survey that found weak support for EOL.[18] Conversely, their larger follow-on study indicated little support for EOL,[16] and 2 other recent studies also found no support for EOL.[11,17] While these contradictory results are surprising, this may in part be due to differences in framing and designs across the surveys, and further research examining this is encouraged. The results suggest some support for BOI, although the findings are not consistent across all models.

Weights

The marginal rates of substitution between BOI and EOL to QALY gain provide evidence for the weights that this research implies. The following example illustrates the effects of the MRS of –0.04 for BOI ($MRS^{BOI}_{(1)}$) and –3.33 for EOL ($MRS^{EOL}_{(1)}$). Let us begin with EOL, which has a discrete effect. Suppose there is a non-EOL patient group with an expected health gain of 5 additional QALYs from new treatment. The societal value of this treatment will be the same as a new treatment for a second group of patients at EOL with an expected QALY gain as small as 1.66. The higher priority afforded by being EOL is exactly cancelled out by a poorer health gain by 3.33 QALYs. On the other hand, the effect of BOI is continuous. Suppose there is a third patient group with BOI of 5 lost QALYs and an expected health gain of 5 additional QALYs from new treatment. If a fourth patient group had an expected health gain as small as 1.66 additional QALYs, then to be equivalent in social value to the third group's treatment, BOI would need to be as high as 83 QALYs. In other words, it takes 83 extra units of BOI to compensate for the lower health gain by 3.33 QALYs. There are very few conditions with a BOI as high as 83, and indeed most have been found to be below 20.

The value of the MRS depends on the size of the QALY gain once the QALY squared term is included, but for most QALY gains, this is unlikely to make much difference. More important, the proportional weight given to each unit of BOI or EOL reduces as the size of QALY gain from new treatment increases.

For example, for $MRS^{BOI}_{(1)}$, 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, BOI:QALY of –0.4:0.05, –0.4:0.5, or –0.4:5. This means that BOI or EOL 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.

Compensating variation is an alternative method for generating QALY weights that is based on standard microeconomic theory.[31] MRS was used rather than compensating variation, as this is the easiest to interpret and most transparent approach that simply reflects the ratio of the coefficients. The DCE was designed to estimate an additive model, as the multiplicative model used elsewhere is not sufficiently transparent enough to be replicated.[32] The use of the additive model means that the results may be affected by a possible interaction between the levels of BOI and QALY gains. Furthermore, weights could theoretically be produced for interventions with positive BOI or EOL even when the QALY gain is zero. However, it is extremely unlikely in a policy context that weights would be required for an intervention providing zero QALY gains.

The conceptual framework focuses on prospective health rather than a lifetime health profile, since the latter would mean reporting age. The inclusion of age in this way means that the results would be unlikely to be used to inform public policy resource allocation; allocation and restriction of health care resources on the basis of age (unless there are clinical reasons to do) are regarded as discriminatory in many countries.[13] The use of prospective health may suggest possible ages, but earlier preparatory studies indicated that age was not an important consideration for respondents completing the types of questions used in our research.[23]

Overall, the results indicated that the coefficients differed across the different variants, where these varied both by life expectancy without the condition in the profiles (N), by the different levels of the attributes, and by the design used to select the profiles (although each design was generated identically using the constraints appropriate for the levels of the attributes in the design). It is not possible to determine which of the differences produced the variation in results. However, it does question the reliability of the use of a single life expectancy, as used in other studies that have followed a similar approach (see, e.g., Shah et al.[16]). Furthermore, this has implications for the pooled models that combine data across all life expectancies without the condition and for the weights that are based on the pooled data.

Limitations

An important concern is the use of an online sample from an existing panel and whether it is representative of the United Kingdom. Online samples may exclude groups in society, such as the computer illiterate or those unable to access the Internet. Members of online panels have stated that they are willing to regularly answer online surveys, making them unrepresentative of computer users. Respondents typically receive points that can be exchanged for goods for every survey they complete, and this may influence the motivation for answering the survey.

Preparatory studies undertaken before the main survey suggested that some respondents failed to understand the concept of BOI in the DCE task, and the main survey was designed to minimize this problem. Respondents' views of the survey indicated that the majority of respondents did not find the DCE tasks difficult, suggesting that respondents felt that they understood the questions that were asked. Robustness analyses indicated that removing respondents who may not have engaged with or understood the survey only affected the coefficients for BOI, and this improved the fit of the model.

Respondents were not allowed to give equal priority to treating both patient groups in the DCE task. However, forced choices for a small group of respondents who are genuinely indifferent should create random noise in the analysis rather than affect the results.

Another possible methodological limitation is the assumption of zero time preference. The scenarios considered in the pairs differ in terms of life expectancy with current and new treatment, and the variants differ in terms of normal life expectancy. For example, a respondent who has a positive time preference would give a lower value to the QALY gain, which has the effect of increasing the MRS.

Normal HRQOL was assumed to be 100%, whereas in reality, it is somewhat below 100%. This assumption was made to simplify the task for respondents, who could then concentrate on other aspects of the task. This assumption needs to be tested in further research using suboptimal normal levels of HRQOL.

The concept of EOL is constrained by the definitions used in the survey. The NICE definition of end of life states under 2 years of life expectancy, whereas the models estimated here use 2 years or less due to the limited number of levels for the life expectancy

variable. Models have been explored elsewhere using life expectancy with current treatment as an explanatory variable rather than the EOL dummy, and the results support the findings reported here.[23] The NICE definition also requires that the condition is a ''rare'' disease,[15] but this has not been considered here.

We should emphasize that this research is not intended to be used at the clinical level. Clinicians cannot be expected to make these types of judgements on an individual level, and indeed they are not permitted to do so (at least within the United Kingdom). The research is intended to provide agencies who use the QALY metric with evidence on the weights they might consider applying to QALY gains to recipients according to their BOI (i.e., QALY loss from the condition), size of QALY gain (to allow for dispersion), and whether they are approaching the end of their life. The calculation of the marginal rates of substitution provides some order of magnitude for the weights that might be considered. This article presents the first attempt to operationalize the concept of BOI, and it provides evidence on societal preferences for EOL. The results indicate general support for maximizing QALY gains; diminishing marginal social value for QALY gains; some support for BOI, although not robust; and robust and consistent support for EOL. The results indicate that a QALY is not a QALY regardless of the burden of the disease or whether the patient is considered end of life and provides a basis for determining appropriate weights.

REFERENCES

1. Claxton K, Martin S, Soares M, et al. Methods for the Estimation of the NICE Cost Effectiveness Threshold. York, UK: University of York; 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;13(5):437–52.

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

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

5. Dolan P, Tsuchiya A. Health priorities and public preferences: the relative importance of past health experience and future health prospects. J Health Econ. 2005;24(4):703–14.

6. Brazier JE, Ratcliffe J, Solomon JA. Measuring and Valuing Health for Economic Evaluation. Oxford, UK: Oxford University Press; 2007.

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

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

9. Nord E. The trade-off between severity of illness and treatment effect in cost-value analysis of health care. Health Policy. 1993; 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;19(4):343–66.

11. Olsen JA. Priority preferences: ''end of life'' does not matter, but total life does. Value Health. 2013;16(6):1063–6.

12. Department of Health. Value Based Pricing: Impact Assessment. London (England): Department of Health (UK); 2010.

13. National Institute for Health and Care Excellence. Consultation Paper: Value Based Assessment of Health Technologies. London: NICE; 2014.

14. van de Wetering EJ, Stolk EA, van Exel NJ, Brouwer WB. Balancing equity and efficiency in the Dutch basic benefits package using the principle of proportional shortfall. Eur J Health Econ. 2013;14(1):107–15.

15. National Institute for Health and Care Excellence. Appraising Life-Extending, End of Life Treatments. London: NICE; 2009.

16. Shah KK, Tsuchiya A, Wailoo AJ. Valuing health at the end of life: a stated preference discrete choice experiment. Soc Sci Med. 2015;124:48–56.

17. Linley WG, Hughes DA. 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 Econ. 2013;22(8):948–64.

18. Shah KK, Tsuchiya A, Wailoo AJ. Valuing health at the end of life: an empirical study of public preferences. Eur J Health Econ. 2014;15(4):389–99.

19. Rowen D, Brazier JE, Mukuria C, et al. Update: Eliciting Societal Preferences for Weighting QALY's According to Burden of Illness, Size of Gain and End of Life. Sheffield, UK: Universities of Sheffield and York: Policy Research Unit in Economic Evaluation of Health and Care Interventions; 2014 Mar 1. Report No.: 018.

20. Rowen D, Brazier JE, Keetharuth A, et al. Comparison of modes of administration and alternative formats for eliciting societal preferences for burden of illness. Appl Health Econ Health Policy. 2015 Oct 8 [Epub ahead of print].

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

22. Kuhfeld WF. Marketing Research Methods in SAS. Cary, NC: SAS Institute; 2005.

23. Brazier JE, Rowen D, Mukuria C, et al. Eliciting Societal Preferences for Burden of Illness, Therapeutic Improvement and End of Life for Value Based Pricing. Sheffield, UK: Universities of Sheffield & York: Policy Research Unit in Economic Evaluation of Health and Care Interventions; 2013 Oct 1. Report No.: 011.

24. McFadden D. Conditional logit analysis of qualitative choice behaviour. In: Zarembka P, ed. Frontiers in Econometrics. New York: Academic Press; 1974. p 105–42.

25. Luce RD. Individual Choice Behaviour: A Theoretical Analysis. New York: 1959.

26. Dolan, P. Modelling valuations for EuroQol health states. Medical Care. 1997;35: 1095–1108.

27. Stata Statistical Software [computer program]. Version Release 11. College Station, TX: StataCorp LP; 2009.

28. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, eds. Second International Symposium on Information Theory. Budapest: Akademiai Kiado, 1973. p 267–81.

29. Schwartz G. Estimating the dimension of a model. Ann Stat. 1978;6:461–4.

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

31. Kind P, Hardman G, Macran S. UK Population Norms for EQ-5D. York, UK: University of York; 1999.

32. Lancsar E, Wildman J, Donaldson C, et al. Deriving distributional weights for QALYs through discrete choice experiments. J Health Econ. 2011;30(2):466–78.

33. Norman R, Hall J, Street D, Viney R. Efficiency and equity: a stated preference approach. Health Econ. 2013;22(5):568–81.

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

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

36. Dolan P, Edlin R, Tsuchiya A. The relative societal value of health gains to different beneficiaries. Health Economics and Decision Science Discussion Paper Series, No 08/12. Sheffield (England): University of Sheffield; 2008. Available from http://eprints.whiterose.ac.uk/10902/1/HEDS_DP_08-12.pdf.