Appl Health Econ Health Policy
DOI 10.1007/s40258-015-0197-y

ORIGINAL RESEARCH ARTICLE

Comparison of Modes of Administration and Alternative Formats for Eliciting Societal Preferences for Burden of Illness

Donna Rowen[1] · John Brazier[1] · Anju Keetharuth[1] · Aki Tsuchiya[1,2] · Clara Mukuria[1]

© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract

Background Proposals for value-based assessment, made by the National Institute of Health and Care Excellence (NICE) in the UK, recommended that burden of illness (BOI) should be used to weight QALY gain. This paper explores some of the methodological issues in eliciting societal preferences for BOI.

Aims This study explores the impact of mode of administration and framing in a survey for eliciting societal preferences for BOI.

Methods A pairwise comparison survey with six arms was conducted online and via face-to-face interviews, involving two different wordings of questions and the inclusion/exclusion of pictures. Respondents were asked which of two patient groups they thought a publically funded health service should treat, where the groups varied by life expectancy without treatment, health-related quality of life (HRQOL) without treatment, survival gain from treatment, and HRQOL gain from treatment. Responses across different modes of administration, wording and use of pictures were compared using chi-squared tests and probit regression analysis controlling for respondent socio-demographic characteristics.

Results The sample contained 371 respondents: 69 were interviewed and 302 completed the questionnaire online. There were some differences in socio-demographic characteristics across the online and interview samples. Online respondents were less likely to choose the group with higher BOI and more likely to treat those with a higher QALY gain, but there were no statistically significant differences by wording or the inclusion of pictures for the majority of questions. Regression analysis confirmed these results. Respondents chose to treat the group with larger treatment gain, but there was little support for treating the group with higher BOI. Respondents also preferred to treat the group with treatment gains in life expectancy rather than HRQOL.

Conclusions Mode of administration did impact on responses, whereas question wording and pictures did not impact on responses, even after controlling for the socio-demographic characteristics of respondents in the regression analysis.

Key Points for Decision Makers

Mode of administration impacted on responses to questions eliciting societal preferences for burden of illness.

In general, question wording and the inclusion of pictures did not impact on responses to questions eliciting societal preferences for burden of illness.

1 Introduction

Economic evaluation is used to determine whether new healthcare interventions are cost effective and hence should be reimbursed. Cost-utility analysis can be used to

✉ John Brazier
j.e.brazier@sheffield.ac.uk

[1] Health Economics and Decision Science, School of Health and Related Research (ScHARR), University of Sheffield, Regent Court, 30 Regent Street, Sheffield S1 4DA, UK

[2] Department of Economics, University of Sheffield, Sheffield, UK

Published online: 08 October 2015

determine whether the intervention is cost effective by generating the cost per quality-adjusted life-year (QALY) gained. The QALY combines length of life with quality of life with the latter based on average individual preferences for different health states.

This study is concerned with the role of mode of administration and alternative formats of questions on responses to questions about the social value of a QALY. One position is to assume that one QALY is always of equal societal worth regardless of who receives the QALY; "a QALY is a QALY" [1], and this is often the implicit basis for cost-utility analyses. However, it has been suggested that some people are more deserving of QALY gains than others, such as on the basis of age, severity of disease or responsibility [2–5], and so their QALYs should be given a higher weight (and hence others a lower weight). Recently, there has been policy interest in the UK in two particular attributes. The National Institute of Health and Care Excellence (NICE) attaches more value to life expectancy gains in patients near to their end of life than those who are not (the 'end of life' criterion [6]), though the weighting it uses is not based on public preferences. More recently, there has been policy interest in a broader notion of severity that captures health-related quality of life (HRQOL) and survival known as 'burden of illness' (BOI) [7, 8]. BOI operationalises the idea of disease severity using the prospective QALY loss suffered by patients with a condition such as breast cancer (allowing for current treatments), compared with their expected QALYs in the absence of the condition, from the decision point relating to a new treatment. As far as we are aware, no other prior study has elicited preferences for this version of severity.

In order to elicit societal preferences for different attributes, there are a number of methodological decisions to be made: (i) elicitation technique, (ii) mode of administration and (iii) survey framing. The method for eliciting preferences selected for this study is a discrete choice experiment (DCE) that asks respondents a series of pairwise comparisons of groups of patients who vary in the attributes of interest. DCE is a well established technique that has started to be used to estimate social value weightings for QALYs [9]. DCE can be used in internet online surveys to provide speedy and comparatively cheap means of collecting data. However, there may be concerns about the quality of the data from poor understanding and engagement by respondents with the task of making social choices.

Two key modes of administration that can be used to collect preference data are face-to-face interviews and internet-based online surveys. Face-to-face interviews have been most widely used to elicit preferences, but in recent years there has been increased interest in collecting preference data online [10–12]. Each mode has advantages and disadvantages that can impact on the data that are generated and therefore the choice of mode needs careful consideration. Face-to-face interviews are expensive and time consuming and as a result often have a lower sample size. The interviewer can monitor understanding, potentially leading to higher quality data and completion rates, yet there are concerns that there may be an interviewer effect where responses are influenced by the presence of the interviewer. Online surveys enable a large number of responses to be collected quickly and relatively cheaply but raise concerns about the quality of the data as without an interviewer present there is no guarantee that the respondent has understood or engaged with the survey. Neither face-to-face interviews nor online surveys typically achieve samples that are fully representative of the general population. Face-to-face interviews usually involve interviewing respondents in their own home and so depend on who is at home at the time the interviewer visits, whereas online surveys are restricted to respondents who have access to the internet, and, if an online panel is used to recruit respondents, then respondents who have self-selected to answer online surveys. This means that mode of administration has important implications for the sample of respondents.

There is a limited literature examining the impact of mode of administration for the elicitation of societal preferences regarding how QALYs are distributed. Two studies examined the use of the person trade-off elicitation technique and compared responses collected online and face-to-face via computer-assisted personal interview [10, 11], finding overall broadly similar responses across the different modes of administration. Another study examined the use of binary choice health state valuation questions online and in a face-to-face computer-assisted personal interview (CAPI), again finding overall similar responses across the different modes of administration [12].

A further issue is the wording of the questionnaire. There is rather less evidence on the role of wording on eliciting societal preferences compared with the broader preferences literature; nonetheless, there has been work on the role of providing clinical information and this was found to have little impact [13–15]. There are many other issues to consider, such as whether the label of 'group of patients' is easier to understand than 'condition group', although the latter is more accurate. Other choices include whether the effects of treatment be worded using either the change in life expectancy or the levels of life expectancy before and after treatment, and whether pictures are used to indicate the size of treatment gains. Framing effects are well acknowledged in the psychology literature but are under-explored in the elicitation of societal preferences. With a complex topic for the survey, it is also important to ensure that questions are worded and presented clearly and simply to enable respondents to understand and engage with the questions.

This study seeks to explore the impact of mode of administration; framing of questionnaire wording (e.g. labelling choice options using a label that says condition or patient group); and use of pictures on a survey, based on a framework designed to elicit societal preferences for BOI, and makes recommendations for future surveys using a similar framework. A pairwise comparison survey with six arms was conducted, where three arms were administered in an online survey and three arms were administered using face-to-face interviews.

2 Methods

2.1 The Framework

The concept of BOI is presented in Fig. 1. Prospective BOI is measured from now, the point at which the healthcare intervention is being considered (for simplification in the survey it is assumed that the patient either has or does not have treatment). From this point, patients have an HRQOL without the treatment, represented as health level H with life expectancy E. Patients given the treatment will gain HRQOL Q and years of survival S. If the patient did not have the condition, they would have HRQOL at 100 % for 20 years. BOI is generated as the difference between the expected prospective health profile without the condition and the prospective health profile with the condition but without the treatment, calculated as $A + B + C + D$. This can be separated into QALY loss due to premature mortality (areas $A + C$ in Fig. 1) and QALY loss from morbidity (areas $B + D$ in Fig. 1). Treatment gain measured in QALYs is equal to the area of C for survival plus D for HRQOL.

2.2 Survey Arms

The same questions (outlined below) were asked using six different arms. The arms were selected to provide evidence to inform the following research questions:

- Do responses differ when the questionnaire is administered online or in a face-to-face interview?
- Do responses differ by the framing used to word the questions?
- Do responses differ by the inclusion of pictures in the questions?

The arms differed by:

- *Mode* two different modes of administration were used: face-to-face interviews and online.
- *Wording* two different wordings of questions were used. One wording involved a condition label and the scenario was described using change in HRQOL and life expectancy due to having the condition and change in HRQOL and life expectancy due to treatment. The other wording involved a patient group label and the scenario was described using the levels of HRQOL and life expectancy without and with treatment. The labels were chosen as the different groups of patients can be

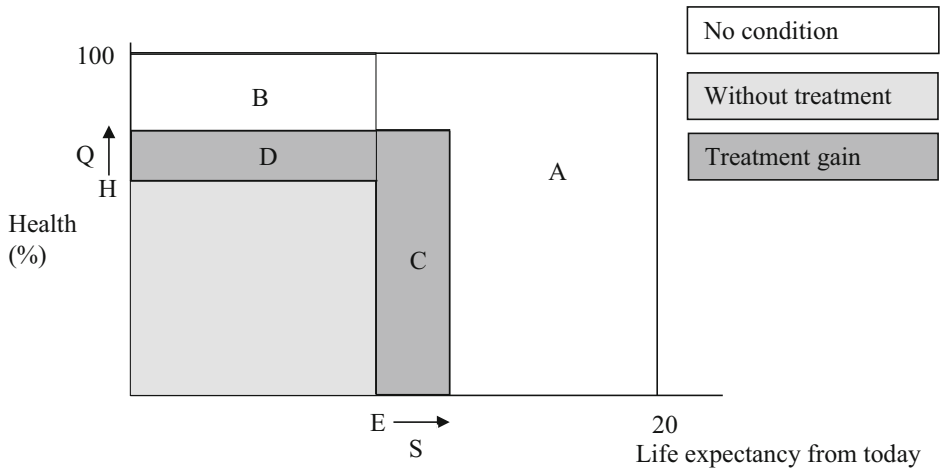


Fig. 1 Representation of profile used in survey. QALY gain $= C + D$ and burden of illness $= A + B + C + D$. *E* life expectancy without treatment, *H* HRQOL without treatment, *HRQOL* health-related quality of life, *Q* HRQOL gain from treatment, *S* survival gain from treatment

thought of as either belonging to different patient groups or conditions, but these terms may not be equally clear or easy to understand for respondents. The two methods of describing the scenario (using change or levels) were chosen as, whilst change is a simpler concept, it does not explicitly provide the information of the level following the change, whereas providing the levels in contrast does not explicitly provide the size of the change. It is unclear which is clearest and easiest to understand for respondents.

- *Video* an introductory video was used to introduce the survey and the concepts used in the survey with the exception of one arm involving face-to-face interview and no pictures, which had no video.
- *Pictures* for the latter wording, one questionnaire framing included pictures, the other included no pictures.

The framing of questions and mode of administration across each arm is summarised in Table 1.

2.3 Pairwise Comparison Questions

This study used pairwise comparison questions to inform the implementation of a DCE. Respondents were asked which group they thought the National Health Service (NHS) should treat, with no indifference option (see "Appendix"). The questions used in this study are not based on a formal statistical design to select the scenarios as in a DCE, and instead, a smaller number of questions were selected (see Table 2) to provide evidence to inform the following research questions:

- Do respondents choose to treat the group with larger QALY gain from treatment? (practice question 1 and question 4),
- Do respondents choose to treat the group with higher BOI? (practice question 2 and questions 1–4),
- Do respondents choose to treat the group with treatment gains in either HRQOL or in life expectancy (where treatment gain measured using QALYs is equal)? (questions 5–7),

The study was reviewed and approved in March 2012 by the School of Health and Related Research (ScHARR) Research Ethics Committee at the University of Sheffield.

2.4 The Survey

2.4.1 Interview Setting

Respondents were sampled to be representative of the UK in terms of age and gender. To try to ensure respondents in each arm had the same characteristics, interviewers selected questionnaires sequentially through the arms to minimise the risk of bias from a correlation of interview location and type of questionnaire. Sampled respondents received a letter and information sheet informing them of the study and that interviewers would be calling at homes in their area. Interviewers knocked on doors and arranged a convenient time to conduct the interview. Respondents were interviewed in their own home by trained and experienced interviewers who had worked on previous valuation surveys conducted by the University of Sheffield.

After giving informed consent, respondents in arms 1 and 3 watched a short video on a laptop explaining the questions. Respondents in arm 2 were read a short introduction by the interviewer with the same general content as the video but excluding the explanation of the pictures since these were not included in this arm. Each arm had two practice questions that involved the interviewer reminding respondents of the factors to consider when making their choice and subsequently an explanation of their choice with one chance for respondents to change their mind and treat the other group. This process was used to enhance respondent understanding of the survey. Following the two practice questions, respondents completed seven interviewer-administered pairwise comparison questions, self-complete questions on attitudes (not reported here), own health in EQ-5D (indicating no problem, some problem or severe problem in mobility, self-care, usual activities, pain or discomfort and anxiety or depression) and socio-demographics and then interviewer-administered questions on their understanding and what they thought of the survey. At the end of the interview, the interviewer reported their perception of respondent understanding, effort and concentration.

2.4.2 Online Setting

Respondents from an online panel were contacted via email to participate in the survey. Respondents were sampled to be representative of the UK in terms of age and gender except for a smaller sample of respondents aged over 65 years. At the start of the survey, respondents were presented with an information sheet and gave informed consent to participate in the survey. Respondents were then shown a short video explaining the questions. It could not be ensured that respondents watched the video, but the video had to be played in full before the respondent could proceed to the practice questions.

As with the interview arms, there were two practice questions which involved an explanation of their choice with a chance for respondents to change their mind. The only differences were that the information could not be read aloud and was instead displayed on the screen; and respondents who changed their mind were then allowed up

Table 1 Summary of the characteristics of each arm of the survey

Arm	Mode of administration	Description of group untreated	Description of treatment	Pictures	Video	Summary
1	Face-to-face interview	Condition *A*/*B*: Reduces health from 100 to H % AND reduces life expectancy from N to E years	Treatment: Restores health from H to $(H + Q)$ % AND extends life expectancy from E to $(E + S)$ years	Yes	Yes	Interview Condition label Wording uses change Pictures Video
2	Face-to-face interview	Patient group *A*/*B* Without treatment: Will live for E years from today With H % health	With treatment: Will live for $(E + S)$ years from today With $(H + Q)$ % health	No	No	Interview Patient group label Wording uses levels No pictures No video
3	Face-to-face interview	Patient group *A*/*B* Without treatment: Will live for E years from today With H % health	With treatment: Will live for $(E + S)$ years from today With $(H + Q)$ % health	Yes	Yes	Interview Patient group label Wording uses levels Pictures Video
4	Online	Condition *A*/*B*: Reduces health from 100 to H % AND reduces life expectancy from N to E years	Treatment: Restores health from H to $(H + Q)$ % AND extends life expectancy from E to $(E + S)$ years	Yes	Yes	Online Condition label Wording uses change Pictures Video
5	Online	Patient group *A*/*B* Without treatment: Will live for E years from today With H % health	With treatment: Will live for $(E + S)$ years from today With $(H + Q)$ % health	No	Yes	Online Patient group label Wording uses levels No pictures Video
6	Online	Patient group *A*/*B* Without treatment: Will live for E years from today With H % health	With treatment: Will live for $(E + S)$ years from today With $(H + Q)$ % health	Yes	Yes	Online Patient group label Wording uses levels Pictures Video

Table 2 Summary of pairwise comparison questions

Q	Choice		Difference between choices	Scenario A				Scenario B			
	A	*B*		*H*	*E*	*Q*	*S*	*H*	*E*	*Q*	*S*
PQ1	Lower treatment gain	Higher treatment gain	BOI same across groups	50	10	10	1	50	10	20	2
PQ2	Higher BOI	Lower BOI	Treatment gain equal across groups[a] Treatment gain only in life expectancy BOI differs only in life expectancy	50	5	0	1	50	10	0	1
1	Higher BOI	Lower BOI	Treatment gain equal across groups[a] Treatment gain only in HRQOL BOI differs only in HRQOL	25	10	10	0	50	10	10	0
2	Higher BOI	Lower BOI	Treatment gain equal across groups[a] Treatment gain only in life expectancy BOI differs in HRQOL and life expectancy	25	5	0	2	50	10	0	1
3	Lower BOI	Higher BOI	Treatment gain equal across groups[a] Treatment gain only in HRQOL BOI differs in HRQOL and life expectancy	50	10	10	0	25	5	20	0
4	Higher BOI	Lower BOI	Treatment gain higher for higher BOI group[ab] Treatment gain in HRQOL and life expectancy BOI differs in HRQOL and life expectancy	25	5	20	2	50	10	10	1
5	Gains in life expectancy	Gains in HRQOL	Treatment gain equal across groups[a] BOI equal across groups HRQOL and life expectancy without treatment at middle levels	50	10	0	2	50	10	10	0
6	Gains in HRQOL	Gains in life expectancy	Treatment gain higher in HRQOL[a] BOI equal across groups HRQOL is high and life expectancy is low without treatment	75	5	20	0	75	5	0	1
7	Gains in HRQOL	Gains in life expectancy	Treatment gain equal across groups[a] BOI equal across groups HRQOL is low and life expectancy is high without treatment	25	15	5	0	25	15	0	3

PQ1 refers to practice question 1 and PQ2 refers to practice question 2

BOI burden of illness, *E* life expectancy without treatment (years), *H* HRQOL without treatment (%), *HRQOL* health-related quality of life, *Q* HRQOL gain from treatment (%), *S* survival gain from treatment (years)

[a] Treatment gain is measured in quality-adjusted life-years (QALYs)

[b] Due to an error, treatment gain was the same for both groups in arm 1

to seven attempts at each practice question before moving on automatically to the next question. Following the two practice questions, respondents completed seven pairwise comparison questions and questions on attitudes, own health in EQ-5D and socio-demographics, understanding and what they thought of the survey. All questions were compulsory.

2.5 Analysis

Responses between arms are compared using the chi-square test to determine whether responses differ by mode of administration (interview versus online, arms 1 and 3 vs arms 4 and 6), the wording of the questions ('condition' label and wording uses change versus 'patient group' label and wording uses levels, arm 4 vs arm 6), and the inclusion of pictures in the questions (pictures versus no pictures, arm 5 vs arm 6). The interview arms (arms 1–3, $n = 23$ each arm) are not powered to assess statistically significant differences between the three of them.

Regression analysis is used to assess, for each of the seven questions, the impact of mode of administration and framing on the pairwise comparison responses whilst controlling for the underlying socio-demographics of respondents. The standard specification for each pairwise comparison question is

$$y* = \alpha + \beta\mathbf{x} + \delta\mathbf{z} + \varepsilon \quad (1)$$

where the dependent variable, $y*$, represents the preference to treat the group with higher BOI for questions 1–4 and the preference to treat the group with gains in life expectancy for questions 5–7 (where BOI is the same across the pairs); $\mathbf{x}$ represents the vector of dummies for mode of administration (arms 1 and 3 vs 4 and 6), wording using levels (4 vs 6), and questions involving pictures and wording using levels (5 vs 6); $\mathbf{z}$ represents the vector of socio-demographic characteristics; and ε represents the error term. As y* is unobserved, all questions are modelled as choices using the probit model and goodness-of-fit statistics are reported. STATA version 11 was used for all regression and statistical analyses.

3 Results

3.1 The Data

Sixty-nine respondents from the North of England were successfully interviewed, providing a response rate of 34 % of respondents who answered the door to the interviewer. Three hundred and two respondents completed the online survey, providing a completion rate of 64 % of people who accessed the survey. All respondents completed every question. No respondents have been excluded from the analysis.

Characteristics of the interview and online samples are summarised in Table 3 along with general population values for England. There were statistically significant differences in age, retired respondents and general health in the samples answering questions involving different modes of administration, and differences in employed/self-employed and retired respondents in the samples answering questions involving different wording and pictures. In comparison with the general population of England, the interview samples were older and contained a larger proportion of retired respondents, whereas the online samples were younger, contained a smaller proportion of employed and retired respondents and were in poorer health.

3.2 Pairwise Comparison Questions

3.2.1 Mode of Administration

There were no statistically significant differences in responses to the practice questions across different modes of administration (see Table 4). However, there were statistically significant differences in responses to questions 3, 4, 5 and 7 across different modes of administration (see Table 5). A larger proportion of respondents in the interviewer arms chose to treat the group with a higher burden of illness (questions 3 and 4) and with higher QALY gain (question 4). A larger proportion of respondents in the online arms chose to treat the group receiving treatment gains in life expectancy (questions 5 and 7).

3.2.2 Wording

There were statistically significant differences in responses to practice question 1 and question 7 across different wordings. A larger proportion of respondents answering questions worded using change and a condition label chose to treat the group with higher treatment gain in practice question 1 in their first response to the question. However, respondents were allowed to change their mind in the practice questions following an explanation of their choice, and once respondents were allowed to change their mind their responses did not differ by wording. A larger proportion of respondents answering questions worded using change and a condition label chose to treat the group with gains in life expectancy in question 7, but no statistically significant difference was observed for the other questions examining a preference for gains in life expectancy rather than gains in health (questions 5 and 6).

Table 3 Sociodemographic characteristics

	Arm						All	England [A]	*p* value mode of administration (arms 1 + 3 vs arms 4 + 6)	*p* value wording (arm 4 vs arm 6)	*p* value pictures (arm 5 vs arm 6)
	1 Interview Condition label Wording uses change Pictures Video	2 Interview Patient group label Wording uses levels No pictures No video	3 Interview Patient group label Wording uses levels Pictures Video	4 Online Condition label Wording uses change Pictures Video	5 Online Patient group label Wording uses levels No pictures Video	6 Online Patient group label Wording uses levels Pictures Video					
N	23	23	23	100	108	94	371				
Mean age (SD)	57.0 (16.6)	50.9 (18.8)	54.5 (17.9)	42.1 (13.7)	40.8 (13.9)	41.7 (15.9)	43.8 (15.8)	NA			
Age distribution (y)									<0.001	0.284	0.628
18–40	8.7 %	34.8 %	26.1 %	46.0 %	48.2 %	50.0 %	43.4 %	41.6 %			
41–65	65.2 %	39.1 %	39.1 %	50.0 %	46.3 %	41.5 %	46.4 %	39.1 %			
Over 65	26.1 %	26.1 %	34.8 %	4.0 %	5.6 %	8.5 %	10.2 %	19.3 %			
Female	56.5 %	69.6 %	47.8 %	51.0 %	51.9 %	54.3 %	53.4 %	51.3 %	0.961	0.650	0.733
Married/Partner	65.2 %	60.9 %	47.8 %	62.0 %	47.2 %	58.5 %	56.1 %	NA	0.638	0.620	0.109
Employed or self-employed	52.2 %	69.6 %	43.5 %	59.0 %	63.0 %	42.6 %	55.3 %	60.9 %	0.696[b]	0.022[b]	0.004[b]
Unemployed	4.4 %	0 %	4.4 %	10.0 %	11.1 %	9.6 %	8.9 %	3.4 %			
Long-term sick	0	0	8.7 %	9.0 %	4.6 %	6.4 %	5.9 %	5.3 %			
Full-time student	4.3 %	0	0	7.0 %	6.5 %	9.6 %	6.5 %	7.3 %			
Retired	39.1 %	26.1 %	43.5 %	9.0 %	7.4 %	20.2 %	16.4 %	13.5 %	<0.001	0.026	0.008
Secondary school is highest level of education	30.4 %	26.1 %	13.0 %	28.0 %	24.1 %	24.5 %	25.1 %		0.524	0.577	0.948
Degree or equivalent professional qualification	26.1 %	52.2 %	43.5 %	51.0 %	46.3 %	44.7 %	46.1 %		0.107	0.379	0.818
Health in general is very good or good	87.0 %	95.7 %	82.6 %	66.0 %	71.3 %	76.6 %	74.4 %		0.059	0.104	0.393
Limited by long-term health condition or disability	17.4 %	26.1 %	26.1 %	39.0 %	34.3 %	25.5 %	31.3 %		0.155	0.045	0.178
EQ-5D score (SD)	0.88 (0.23)	0.91 (0.12)	0.84 (0.27)[B]	0.78 (0.27)	0.81 (0.24)	0.81 (0.24)	0.82 (0.24)	0.86 (0.23)[C]	0.126[c]	0.366[c]	0.890[c]

NA not available, *SD* standard deviation

[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 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

[B] One missing observation

[C] Interviews conducted in the Measurement and Valuation of Health (MVH) study [6]

[b] This has been calculated for employed or not employed rather than employment status due to the small number of observations in several employment status categories

[c] A *t* test has been used to test statistical significance

Table 4 Responses to the practice questions

Q	Choice	Explanation	Arm						*p* value mode of administration (arms 1–3 vs arms 4–6)	*p* value wording (arm 4 vs arms 5 and 6)	*p* value pictures (arm 5 vs arm 6)
			1 Interview Condition label Wording uses change Pictures Video ($N = 23$) (%)	2 Interview Patient group label Wording uses levels No pictures No video ($N = 23$) (%)	3 Interview Patient group label Wording uses levels Pictures Video ($N = 23$) (%)	4 Online Condition label Wording uses change Pictures Video ($N = 100$) (%)	5 Online Patient group label Wording uses levels No pictures Video ($N = 108$) (%)	6 Online Patient group label Wording uses levels Pictures Video ($N = 94$) (%)			
PQ1 First response	A		4.3	21.7	4.3	7.0	5.6	16.0	0.155	0.049	0.016
	B	Higher gain	95.7	78.3	95.7	93.0	94.4	84.0			
PQ1 Final response	A		4.3	21.7	4.3	7.0	7.4	11.7	0.277	0.259	0.297
	B	Higher gain	95.7	78.3	95.7	93.0	92.6	88.3			
PQ2 First response	A	Higher BOI	39.1	56.5	56.5	53.0	50.9	41.5	0.961	0.109	0.180
	B		60.9	43.5	43.5	47.0	49.1	58.5			
PQ2 Final response	A	Higher BOI	39.1	60.9	56.5	56.0	54.6	44.7	0.743	0.115	0.158
	B		60.9	39.1	43.5	44.0	45.4	55.3			

PQ1 refers to practice question 1 and PQ2 refers to practice question 2

In Arms 1–3, interviewers recorded first choice and then whether the respondent changed their mind following the explanation of their choice

In Arms 4–7 in the online survey, respondents who stated after the explanation of their choice that they did not still want to treat the same group were asked the question again, and this process was repeated up to seven times before moving on to the next question

BOI burden of illness

Table 5 Responses to the pairwise comparison questions

Q.	Choice	Explanation	Arm						*p* value mode of administration (arms 1–3 vs arms 4–6)	*p* value wording (arm 4 vs arms 5 and 6)	*p* value pictures (arm 5 vs arm 6)
			1 Interview Condition label Wording uses change Pictures Video ($N = 23$) (%)	2 Interview Patient group label Wording uses levels No pictures No video ($N = 23$) (%)	3 Interview Patient group label Wording uses levels Pictures Video ($N = 23$) (%)	4 Online Condition label Wording uses change Pictures Video ($N = 100$) (%)	5 Online Patient group label Wording uses levels No pictures Video ($N = 108$) (%)	6 Online Patient group label Wording uses levels Pictures Video ($N = 94$) (%)			
Q1	A	Higher BOI	39.10	56.50	52.20	51.00	47.20	44.70	0.780	0.379	0.718
	B		60.90	43.50	47.80	49.00	52.80	55.30			
Q2	A	Higher BOI	39.10	43.50	56.50	53.00	46.30	46.80	0.791	0.389	0.942
	B		60.90	56.50	43.50	47.00	53.70	53.20			
Q3	A		34.80	34.80	26.10	50.00	38.90	40.40	0.066	0.181	0.824
	B	Higher BOI	65.20	65.20	73.90	50.00	61.10	59.60			
Q4	A	Higher BOI and higher gain	52.20[a]	78.30	87.00	62.00	62.00	60.60	0.023[a]	0.846	0.839
	B		47.80	21.70	13.00	38.00	38.00	39.40			
Q5	A	Gains in life expectancy	30.40	69.60	47.80	68.00	69.40	67.00	<0.001	0.884	0.712
	B	Gains in HRQOL	69.60	30.40	52.20	32.00	30.60	33.00			
Q6	A	Gains in HRQOL and higher gain	65.20	47.80	56.50	50.00	57.40	58.50	0.408	0.234	0.874
	B	Gains in life expectancy	34.80	52.20	43.50	50.00	42.60	41.50			
Q7	A	Gains in HRQOL	52.20	43.50	56.50	29.10	37.00	43.60	0.023	0.034	0.341
	B	Gains in life expectancy	47.80	56.50	43.50	71.00	63.00	56.40			

BOI burden of illness, *HRQOL* health-related quality of life

[a] Due to an error, treatment gain was the same for both groups in arm 1, and hence arm 1 has been excluded from the generation of the *p* value for mode of administration

3.2.3 Pictures

The only statistically significant difference across the inclusion of pictures was for practice question 1 for the first response, where a smaller proportion of respondents answering questions including pictures chose to treat the group with higher treatment gain. However, for the final response to the practice question where respondents were allowed to change their mind responses did not differ by the inclusion of pictures.

3.2.4 Treatment Gain

Respondents overwhelmingly chose to treat the group with the highest treatment gain (practice question 1 and question 4).

3.2.5 Burden of Illness

The results vary by question, suggesting that the preferences for BOI are sensitive to the characteristics of the scenario. For three questions there is no clear preference for treating the group with higher BOI (practice question 2 and questions 1 and 2) and for one question there is a general preference for treating the group with higher BOI (question 3). A larger number of respondents chose to treat the group with the higher BOI and higher treatment gain (question 4).

3.2.6 Gains in HRQOL or Life Expectancy

Respondents in arm 2 and the online arms (arms 4–6) have a preference for treating the group with treatment gains in life expectancy, whereas there is no clear preference for arms 1 and 3 (questions 5 and 7). There is no preference for treating the group with gains in life expectancy where the treatment gain is higher in HRQOL (question 6).

3.3 Regression Analyses

Across the seven regression models, the dummy variable representing the online mode of administration was significant for four questions (questions 3, 4, 5 and 7, Table 6). Online respondents were less likely to treat the

Table 6 Regression analyses for the pairwise comparison questions

Pairwise comparison question	1	2	3	4	5	6	7
Dependent variable	Choose to treat group with higher BOI				Choose to treat group with gains in life expectancy		
Online	−0.133	−0.026	−0.528***	−0.696***	0.464**	0.022	0.400**
Wording uses levels (arms 2, 3, 5 and 6)	−0.060	−0.146	0.205	−0.047	0.262	−0.091	−0.141
Wording uses levels and questions involve pictures (arms 3 and 6)	−0.029	0.097	0.073	0.030	−0.195	−0.057	−0.193
Female	0.012	−0.096	−0.024	0.059	−0.123	−0.197	−0.368***
Married	0.020	0.183	0.087	0.095	0.191	0.096	−0.064
Unemployed	−0.008	0.004	0.215	−0.173	0.135	−0.374	0.333
Sick	−0.163	−0.081	−0.107	0.020	−0.136	−0.408	0.469
Student	−0.168	−0.167	0.347	0.195	0.207	0.160	−0.347
Retired	−0.447	−0.303	−0.557*	−0.421	0.205	−0.167	−0.327
Limited by long-term health condition or disability	0.115	0.178	0.267	−0.060	0.092	−0.028	−0.145
Secondary school is highest level of education	−0.443***	−0.251	−0.538***	−0.320*	0.087	0.244	0.215
Aged 41–65 years	−0.354**	−0.380**	−0.350**	−0.404**	−0.126	−0.290*	0.064
Aged over 65 years	0.054	−0.016	−0.107	0.068	−0.203	0.091	0.571
Constant	0.412*	0.242	0.762***	1.279***	−0.166	0.115	0.279
Observations	371	371	371	348	371	371	371
Log likelihood	−245.3	−248.2	−229.9	−211.1	−233.0	−247.0	−234.6
Pseudo *R*-squared	0.0449	0.0339	0.0842	0.0686	0.0329	0.0309	0.0551

Online is a dummy variable representing the online mode of administration. 'Wording uses levels' is a dummy variable that equals 1 if the wording of the questions uses levels and 0 otherwise (equals 1 for arms 2, 3, 5 and 6, equals 0 for arms 1 and 4). 'Wording uses levels and questions involve pictures' is a dummy variable that equals 1 if the wording of the questions uses levels and the questions also involve pictures and 0 otherwise (equals 1 for arms 3 and 6, equals 0 for arms 1, 2, 4 and 5)

BOI burden of illness; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

group with higher BOI and were more likely to choose to treat the group with gains in life expectancy. The dummy variables for 'wording uses levels', and for 'wording uses levels and questions involve pictures' were not significant, suggesting that the framing used here did not have a systematic impact on the pairwise comparison results, consistent with the results outlined above. Few socio-demographic variables were significant and none were consistently significant across all questions.

3.4 Respondent Views of the Survey

Overall, respondents reported that the questions were largely easy to understand, the interviewer explanation was clear, the wording of the questions was clear, the video was useful and that pictures were helpful (Table 7). The proportion of respondents reporting that the wording of the questions was clear was larger in the online arms and the difference was statistically significant, but there were no differences in the difficulty of questions, usefulness of the video or helpfulness of the pictures by mode of administration. There were no differences in respondent views of the survey by wording. The proportion of respondents reporting that the questions were easy to understand and that the video was not useful was larger in the arm that did not include pictures and the difference was statistically significant, but there were no differences in reporting that the wording of the questions was clear by the inclusion of pictures (Table 7).

4 Discussion

Overall, there were differences in responses to the pairwise comparison questions between interviewer and online modes of administration, but there were few differences across different framings using different wordings and the inclusion/exclusion of pictures. Regression analyses controlling for the socio-demographic characteristics of respondents support these results. The findings are contrary to existing studies finding that responses were overall similar across the different modes of administration [10–12]. However, all these studies involved the use of a computer throughout the interviews, whereas our study used paper-based questionnaires in the interviews, with a computer being used only for the introductory video. The use of computers throughout the interview may not enable the same interviewer interaction as face-to-face interviews, and hence for our study this interaction may have impacted on responses.

The results suggest that the different wordings and inclusion/exclusion of pictures used here did not impact on responses, although a higher proportion of respondents in the arm without pictures found the questions easier to understand. Mode of administration did impact on responses, but it cannot be judged which is most preferable. In terms of respondent views of the survey, a higher proportion of respondents in the online arms reported that the questions were clear, although both modes appear feasible.

Overall, the results indicated support for choosing to treat the group with the larger treatment gain, with little support for treating the group with higher BOI, but this was dependent upon the characteristics of the scenarios in each question. Other studies have found a preference for treating patients who have less severe health problems (e.g. [9]), suggesting that a lack of support for BOI may not be surprising. Overall, there was a preference for treating the group with treatment gains in life expectancy rather than treatment gains in HRQOL, contrary to a study on the end-of-life criterion that found greater preferences for HRQOL improvement over life expectancy extension [16].

The key limitation of this study surrounds the extent to which respondents understood the task they were doing. Within the design of this study it was not possible to verify the actual understanding of respondents. Their answers to the question about their understanding were positive in most cases, but this has not been validated here. Most respondents answered the first practice question in a logically consistent way. Nonetheless it is a complex task to choose between groups of patients varying in QALY gains and BOI and we cannot rule out a degree of misunderstanding that may invalidate the results surrounding the importance of QALY gain and BOI. This is likely to be less important for the comparison between modes of administration and ways of framing.

Another important limitation of this survey is the small sample size for each of the interview arms. To account for this, the differences in the interview arms have not been assessed statistically, but this means that the impact of the wording and the inclusion of pictures have not been assessed in the interview arms. In addition, the larger relative size of the online data in the regression analyses may have impacted on the results.

Following this study, a large online study using the question framing of arm 6 was undertaken [17]. An online study was selected for its advantages in terms of time and cost. Whilst arm 5 with no pictures had a larger proportion of respondents finding the questions easier to understand, 86.2 % of respondents in arm 6 felt that the pictures were helpful and for this reason arm 6, which included pictures, was selected. Furthermore, the regression analyses indicated no statistically significant differences in responses to the pairwise questions by questionnaire framing. This larger study also found that respondents chose to treat the group with the larger treatment gain, but found there was some evidence to suggest that respondents chose to treat the group with a larger BOI [17].

Table 7 Respondents views of the survey

	Arm						*p* value mode of administration (arms 1–3 vs arms 4–6)	*p* value wording (arm 4 vs arms 5 and 6)	*p* value pictures (arm 5 vs arm 6)
	1 Interview Condition label Wording uses change Pictures Video (%)	2 Interview Patient group label Wording uses levels No pictures No video (%)	3 Interview Patient group label Wording uses levels Pictures Video (%)	4 Online Condition label Wording uses change Pictures Video (%)	5 Online Patient group label Wording uses levels No pictures Video (%)	6 Online Patient group label Wording uses levels Pictures Video (%)			
Pairwise comparison questions were							0.800	0.579	0.084
Very difficult to understand	0	0	0	1.0	0.9	1.1			
Quite difficult to understand	8.7	13.0	8.7	10.0	4.6	5.3			
Neither difficult nor easy to understand	8.7	8.7	21.7	19.0	19.4	24.5			
Fairly easy to understand	60.9	47.8	39.1	41.0	31.5	45.7			
Very easy to understand	21.7	30.4	30.4	29.0	40.7	23.4			
Interviewer explanation was clear	95.7	100.0	100.0	NA	NA	NA	NA	NA	NA
Wording of the questions was clear	91.3	100.0	100.0	100.0	97.2	100.0	0.004	NA	0.104
Video was useful	78.3	NA	82.6	86.0	69.4	83.0	0.498	0.561	0.025
Pictures were helpful	87.0	NA	78.3	89.0	NA	86.2	0.368	0.550	NA

NA not available

5 Conclusion

This study shows the importance of survey framing and mode of administration. The results indicate that mode of administration did impact on responses, but that framing in terms of question wording and inclusion of pictures had little impact on responses and had no impact after controlling for the socio-demographic characteristics of the samples.

Acknowledgments We are grateful to the survey respondents. We are grateful to Gavin Roberts and Danny Palnoch from DH for their comments throughout this project. We are also especially grateful to Angela Robinson for her advice on the development of the survey and for her suggestions on a previous draft of the paper.

Author contributions Donna Rowen coordinated the study and contributed to the study design, undertook the analyses and wrote a draft of the paper. John Brazier is the PI for the grant, contributed to design, analyses and writing of the paper, and is guarantor of the overall content. Anju Keetharuth and Aki Tsuchiya contributed to the design, analyses and writing of the paper. Clara Mukuria contributed to the analyses and writing of the paper.

Compliance with Ethical Standards

This is an independent report commissioned and funded 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. Donna Rowen, John Brazier, Anju Keetharuth, Aki Tsuchiya and Clara Mukuria have no conflicts of interest. Ethical approval was provided by the ScHARR Ethics Committee in accordance with the Declaration of Helsinki. Informed consent was obtained from participants in the study.

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

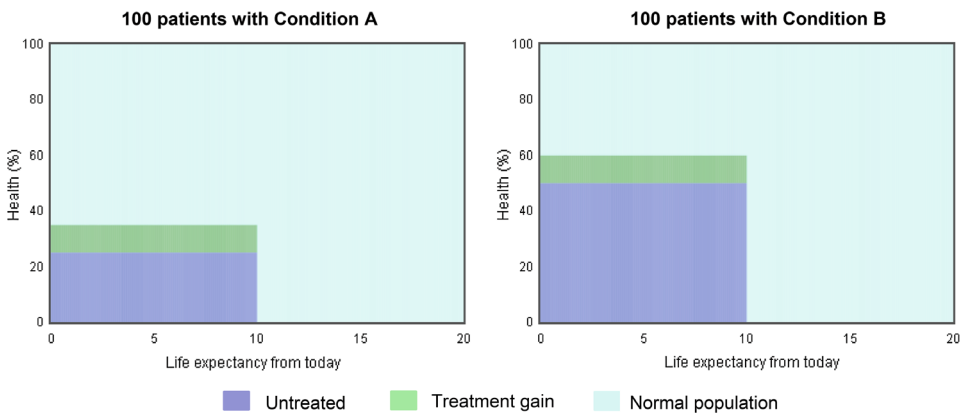
Arms 1 and 4 Question 1

Condition A	**Condition B**
• reduces health from 100% to 25% • AND reduces life expectancy from 20 years to 10 years	• reduces health from 100% to 50% • AND reduces life expectancy from 20 years to 10 years
Treatment • restores health from 25% to 35% • BUT does not extend life expectancy	**Treatment** • restores health from 50% to 60% • BUT does not extend life expectancy

Which group do you think the NHS should treat?

100 patients with Condition A or 100 patients with Condition B?

Condition A	Condition B



Arms 2 and 4 Question 1

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 **25%** health **With treatment** • will live for **10** years from today • with **35%** health	**Patient group B** **Without treatment** • will live for **10** years from today • with **50%** health **With treatment** • will live for **10** years from today • with **60%** health
If these patients did not have a medical condition they would live in 100% health for 20 years from today	

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?

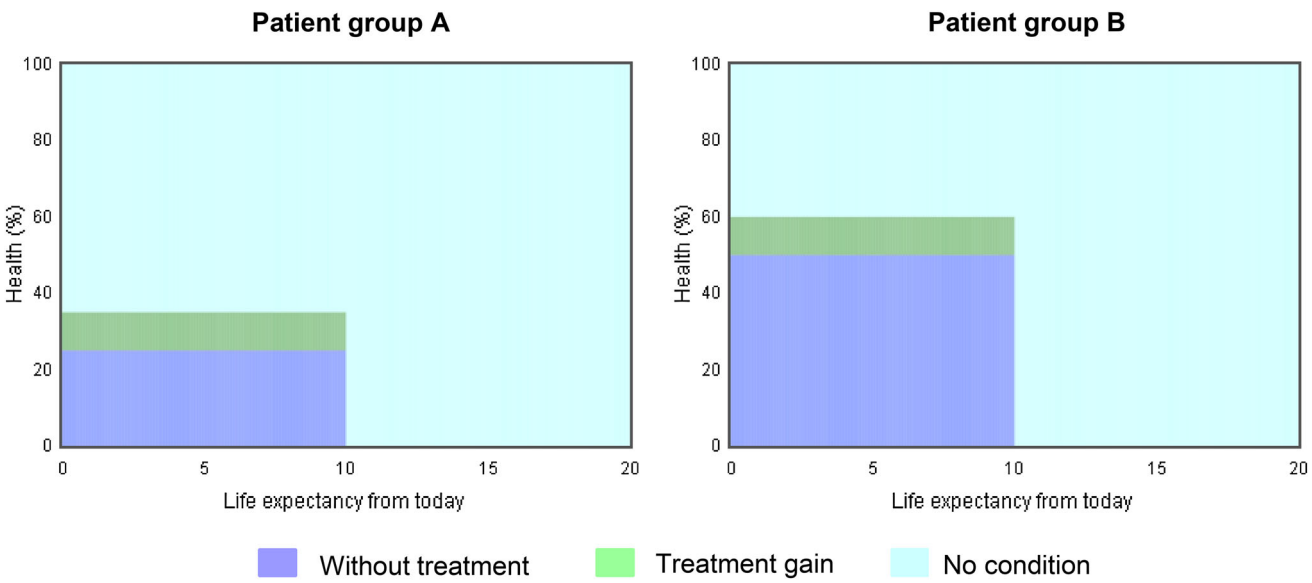
Patient group A	Patient group B

Arms 3 and 6 Question 1

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 **25%** health **With treatment** • will live for **10** years from today • with **35%** health	**Patient group B** **Without treatment** • will live for **10** years from today • with **50%** health **With treatment** • will live for **10** years from today • with **60%** health
If these patients did not have a medical condition they would live in 100% health for 20 years from today	

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



Reference

1. Weinstein MC. A QALY is a QALY- or is it? J Health Econ. 1988;7:289–90.
2. Green C. Investigating public preferences on 'severity of health' as a relevant condition for setting health care priorities. Soc Sci Med. 2009;68(2247):2255.
3. Shah K. Severity of illness and priority setting in healthcare: a review of the literature. Health Policy. 2009;93:77–84.
4. Brazier JE, Ratcliffe J, Solomon JA, Tsuchiya A. Measuring and valuing health for economic evaluation. Oxford: Oxford University Press; 2007.
5. Nord E. The trade-off between the severity of illness and treatment effects in cost-value analysis of health care. Health Policy. 1993;24:227–38.
6. National Institute of Health and Care Excellence (NICE). Appraising life-extending, end of life treatments. London: NICE; 2009 **(ref type: report)**.
7. Department of Health. Value based pricing: impact assessment; 2010 (ref type: report).
8. National Institute of Health and Care Excellence (NICE). Consultation paper: value based assessment of health technologies. 2014. London: NICE. Available at https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/VBA-TA-Methods-Guide-for-Consultation.pdf (ref type: report). Accessed Dec 2014.
9. Lancsar E, Wildman J, Donaldson C, Ryan M, Baker R. Deriving distributional weights for QALYs through discrete choice experiments. J Health Econ. 2011;30:466–78.
10. Covey J, Robinson A, Jones-Lee M, Loomes G. Responsibility, scale and valuation of rail safety. J Risk Uncertain. 2010;40:85–108.
11. Damschroder LJ, Baron J, Hershey JC. The validity of person trade-off methods: randomized trial of computer elicitation versus face-to-face interview. Med Decis Mak. 2004;24:170–80.
12. Mulhern B, Longworth L, Brazier J, Rowen D, Bansback N, Devlin N, Tsuchiya A. Binary choice health state valuation and mode of administration: Head to head comparison of online and CAPI. Value Health. 2013;16:104–13.
13. Roberts T, Bryan S, Heginbotham C, McCallum A. Public involvement in health care priority setting: an economic perspective. Health Expect. 1999;2:235–44.
14. Bryan S, Roberts T. Public involvement in health care priority setting: an economic perspective. Health Expect. 1999;2(4):235–44.
15. Bryan S, Roberts T. QALY-maximisation and public preferences: results from a general population survey. Health Econ. 2002;11(8):679–93.
16. Shah KK, Tsuchiya A, Wailoo A. Valuing health at the end of life: an empirical study of public preferences. Eur J Health Econ. 2014;15:389–99.
17. Rowen D, Brazier J, Mukuria C, Keetharuth A, Risa A, Tsuchiya A, Whyte S, Shackley P. Eliciting societal preferences for weighting QALYs for burden of illness and end of life. Med Decis Mak **(forthcoming)**.