Article:
Rowen, D., Brazier, J., Tsuchiya, A. et al. (2 more authors) (2010) It's all in the name, or is it? The impact of labelling on health state values. HEDS Discussion Paper, 10/09.

HEDS Discussion Paper 10/09

HEDS Discussion Paper 10/09

Published paper

Rowen, D, Brazier, JE, Tsuchiya A, Young, T, Ibbotson, R. It's all in the name, or is it? The impact of labelling on health state values. Medical Decision Making 2012, 32(1), 31-40
www.ncbi.nlm.nih.gov/pubmed/21685376


ScHARR
School of Health and
Related Research

It's all in the name, or is it? The impact of labelling on health state values

Rowen D*[1], Brazier J[1], Tsuchiya A[1,2], Young T[1], Ibbotson R[3]

1. School of Health and Related Research (ScHARR), University of Sheffield
2. Department of Economics, University of Sheffield
3. Centre for Health and Social Care Research, Sheffield Hallam University, UK

* Correspondence to: Donna Rowen, Health Economics and Decision Science, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK.
Telephone: +44114 222 0728.
Fax: +44114 272 4095. Email:
d.rowen@sheffield.ac.uk

3 August 2010

Abstract

Many descriptions of health used in vignettes and condition-specific measures refer to the medical condition. This paper assesses the impact of referring to the medical condition in the descriptions of health states valued by members of the general population. A sample of 241 members of the UK general population each valued 8 health states using time trade-off. All respondents valued essentially the same health states, but for each respondent the descriptions featured either an irritable bowel syndrome label, a cancer label or no label. Regression techniques were used to estimate the impact of each label and experience of the condition on health state values. We find that the inclusion of a cancer label in health state descriptions affects health state values and that the impact is dependent upon the severity of the state. A condition label can affect health state values, but this is dependent upon the specific condition and severity. It is recommended to avoid condition labels in health state descriptions (where possible) to ensure that values are not affected by prior knowledge or preconception of the condition that may distort the health state being valued.

* * * * *

Acknowledgements: We would like to thank all the interviewees who took part in the survey and to participants at Health Economics Study Group Meeting in June 2010 for their comments on an earlier version of this paper. This study was funded by MRC-NIHR Methodology Research Programme, project number 06/97/04.

Introduction

During the past decade resource allocation across competing health care interventions has been increasingly informed by economic evaluation using cost-utility analysis. Cost-utility analysis measures the benefits of competing health care interventions using Quality Adjusted Life Years (QALYs) which capture both changes in quality and quantity of life. Typically the 'Q' quality adjustment weight used to produce the QALY estimate is obtained using an off-the-shelf generic preference-based measure of health such as the EQ-5D (1), which is the most commonly used measure and is recommended by the National Institute of Health and Clinical Excellence (NICE) in England and Wales (2). Whilst the EQ-5D is found to be reliable, valid and responsive for many conditions, such as rheumatoid arthritis (3), for other conditions this is not the case, such as visual impairment (4) and hearing loss (5). For this reason condition-specific measures (CSM) are included in many clinical trials, either alongside or instead of generic preference-based measures. Condition- specific measures will continue to be an important potential source of evidence that can be used to populate cost-effectiveness models for economic evaluation. However the use of condition-specific measures in economic evaluation is currently limited because most condition-specific measures are not 'preference-based' and thus cannot be used to derive the 'quality adjustment weight' used to estimate QALYs.

Several techniques are currently used to obtain estimates of the quality adjustment weight for use in QALYs from existing condition-specific measures: mapping from a condition-specific non-preference based measure to a generic preference-based measure; deriving vignettes and valuing these; deriving a condition-specific preference-based measure from the existing non-preference-based measure. Mapping has gained increasing popularity in recent years for use in cost-effectiveness analysis as it enables a utility estimate to be produced even when a preference-based measure was not used in the trial. Mapping relies upon a large degree of overlap between both measures, yet this is not always the case for condition-specific and generic measures. A recent review of mapping studies found that some studies had large errors in predicted utility values, and that this was more pronounced when mapping a condition-specific measure onto a generic measure (6). Arguably the advantages that condition-specific measures bring through their focus, depth, relevance and sensitivity are lost through mapping to a generic measure. Furthermore, if a generic measure is inappropriate for a condition or population, mapping to a generic measure will also be inappropriate.

Another approach commonly used to populate economic models is to develop vignettes or scenarios describing various health states (for example (7;8)). These can be based on standardised measures or more commonly are 'bespoke' descriptions that are developed specific to the condition and/or treatments being considered. The use of vignettes requires common health states for the trial to be developed and valued, meaning they are resource-intensive and usually trial-specific. This reduces comparability between different trials or interventions when these values are used in economic evaluation. Furthermore, vignettes tend to be developed from expert opinion rather than directly from evidence. Nonetheless this approach continues to

be used in the absence of better data or where a generic measure is regarded as inappropriate.

Preference-based measures have been derived from existing condition-specific measures for a wide variety of conditions, including asthma (9;10) and overactive bladder (11;12). Contrary to vignettes, these have the advantage that they can be used in all trials that include the condition-specific measure, thus enabling comparability within a condition or patient population. A further advantage is that they retain the focus, depth, relevance and sensitivity of the original measure.

One fundamental concern with condition-specific measures is their lack of comparability and applicability across different conditions or patient groups for use in economic evaluation (13). Typically condition-specific measures are used as a supplement rather than the preferred alternative to generic measures for use in economic evaluation. Recent NICE guidance for economic evaluation (2) states that condition-specific measures may be used when EQ-5D data is either inappropriate or unavailable (although in the latter case mapping is preferred). Comparability with EQ-5D can be maintained by using the same valuation 'protocol' including the valuation technique (and variant), procedures, common anchors (full health and dead), visual aids and the same type of respondents (such as a representative sample of the general population). An important difference between generic and condition- specific measures is the general inability of condition-specific measures to capture many co-morbidities and side-effects, although this is specific to the measure, condition, treatment and patient population. The impact of excluding key dimensions is assessed elsewhere (14).

Another important difference between generic and condition-specific measures is that many condition-specific preference-based measures state the cause of the health problems being assessed in the classification system. Non-preference-based condition-specific measures typically name the condition to enable greater precision for assessing changes in quality of life due to that condition and the relevant intervention, which has a significant advantage, for example, for use in drug labelling claims. Often the condition is embedded within the classification system derived from the non-preference-based measure, meaning the system cannot be valued without labelling the condition. For example, for an asthma-specific measure (9;10) it is infeasible to remove the mention of asthma without changing the meaning of the dimension 'experience asthma symptoms as a result of air pollution'. Yet the inclusion of a condition label in a health state valuation exercise may impact on the values provided by members of the general population due to prior knowledge or preconception of the condition. This is problematic as it may mean that the elicited values do not reflect the health state being valued.

There is disagreement in the literature regarding the extent to which naming the condition in health state descriptions has an impact on health state values. Some studies have found that the inclusion of condition labels has lowered health state values (15-19). For example, one study found that the explicit use of mental health labels including mental handicap, schizophrenia and dementia was associated with lower health state values (17). Two studies found that labelling breast cancer states

did not affect values (15;17) with the exception of scenarios written in the third person (15), whereas another study found that it reduced values (18). The finding that cancer labels have no impact is surprising given that the large majority of members of the general population will have prior knowledge and preconceptions of cancer: cancer treatments can have severe treatment side effects associated with low quality of life and cancer is widely known as one of the world's largest killers, especially in developed countries.

The majority of studies examining the impact of labelling include more than one condition label and find that the results differ according to the specific condition. Typically the studies have a small sample size (for example (18) has 26 respondents and (17) has 42 respondents) meaning that it is difficult to test statistical significance, and ask respondents to value states using a large number of different condition labels (for example in (18) one group value states across 9 conditions and the other control group value states with no condition label). All studies ask respondents to value states with changing descriptions (due to framing or labelling) and assess whether values change accordingly. This study design means that there may be a focussing effect, as respondents will realise that the health states are different in part due to the presence of the (changing) condition label (or framing). This may actually cause respondents to change their values accordingly, and to purposefully consider their prior knowledge and preconceptions of the condition. Furthermore the health states presented do not cover a wide range of severity meaning that the results are specific to the small number of states presented. To address these limitations we have undertaken a study to compare values from samples of respondents who value states with only one or no condition label across a range of health states of differing severity.

This paper assesses the impact of referring to the medical condition in the descriptions of health states valued by members of the general population. We compare health state values that were elicited using three different labels: no label; irritable bowel syndrome label; cancer label. Each respondent values the same set of states using only one label. We further analyse values to determine whether values are affected by prior experience of the relevant condition and other background characteristics along with severity of the state.

The classification system from EORTC-8D, a recently developed preference-based measure for cancer derived from the EORTC QLQ-C30 is used to define the health states (20). The use of this classification system rather than vignettes has the advantage that the health state descriptions are based on the EORTC QLQ-C30, which is a widely used, valid and reliable measure of health related quality of life. Despite being developed for cancer the content of the EORTC-8D classification system is quite generic and makes no reference to cancer. A member of the general population (or anyone else) would not necessarily know it was a cancer measure. The classification system has no labels and hence labels can be added and removed without fundamentally changing the meaning of any of the dimensions. The classification system, shown in Table 1, has 8 dimensions: physical functioning; role functioning; pain; emotional functioning; social functioning; fatigue and sleep disturbance; nausea; constipation and diarrhoea. Each dimension has 4 or 5 severity

levels and the classification system defines 81,920 health states. The valuation study used to elicit preference weights for all states defined by the classification system did not include a condition label (20).

Methods

Given that the classification system analysed here is intended for use to capture quality of life changes in cancer patients, and did not include a condition label in the original valuation study, two of the labels chosen for this study were no label and a cancer label. We were further interested in testing the generalisability of this finding to another condition and selected one other condition that might be reasonably perceived by respondents to account for health states described using EORTC-8D. The third label chosen was irritable bowel syndrome (IBS) as several clinicians and doctors suggested that this is a condition that would be accurately described by the EORTC-8D classification system.

Valuation survey

A valuation study was conducted where members of the general population each valued the same 8 health states from EORTC-8D using time trade-off (TTO). The sample was divided into 3 groups: no label, IBS label and cancer label. These groups were interviewed in sequence, with all of the respondents from the first group being interviewed before respondents from the second group were contacted. This was undertaken in order to avoid the risk of contamination.

A representative sample of the general population was used for each of the 3 groups. To ensure each group had the same characteristics and were a representative sample of the UK, the sampling strategy involved two steps. Firstly all households contacted to request participation in the study were sampled using the AFD Names and Numbers version 3.1.25 database (AFD Software Limited, Ramsey, UK) and the sample was then balanced to the UK population according to ACORN profiles. Secondly each unique postcode included in the sample was divided into 3 to form 3 separate samples for each group. The sample size was chosen to enable the comparison of mean values between health states with the inclusion of each condition label using simple t-tests. Assuming a power of 0.8, significance level of 0.05, standard deviation of 0.3 and an expected difference of 0.1, this requires a sample of 73 valuations for each state and a total of 219 completed interviews.

Eight health states were selected to represent the range of possible values for the measure, and this was informed by the results of the original valuation study (20). The original study valued 85 health states across 12 combinations of states (card blocs) each with 8 states, where 81 states were chosen using an orthogonal array and this was supplemented by 4 additional states. Selection criteria for the most appropriate combination of 8 health states (card bloc) to be selected from those valued in the original study included: minimum prediction error per card bloc, largest range of mean TTO distribution per card bloc, smallest missing data per card bloc, and general 'feasibility' of states (for example avoiding 'You were not limited in pursuing your hobbies or other leisure time activities' combined with 'Your physical

condition or medical treatment interfered very much with your social activities'). The selected states are: 11111111, 31212241, 13423411, 44321321, 23141224, 24432411, 51224434, 544444444, where 1 is no problems and level 4 is the worst level (5 for physical functioning dimension). An example health state (51224434) using each label is shown in Figure 2.

At the start of the interview respondents were asked to read and self-complete the classification system with the relevant condition label, to familiarise themselves with the system. Secondly, as a warm-up task prior to TTO respondents were asked to rank 8 health states alongside 'full health' and 'dead'. Thirdly respondents valued the 8 states using the Measurement and Valuation of Health (MVH) study version of TTO including a visual prop designed by the MVH Group (University of York) (1). To familiarize respondents with the TTO task, respondents completed a practice TTO question for an additional state (22332322) prior to valuing the 8 health states. Finally respondents self-completed questions covering their demographic and socio-economic characteristics.

For the two groups involving a label, the condition was mentioned in the cover letter, information sheet about the study, and on the cards used in the valuation exercises. Prior to self-completing the EORTC-8D classification system respondents in the labelling groups were shown an information sheet about the relevant condition (included in the appendix). A question on whether the respondent has experience of the condition, in themselves, in family or in caring for others, was included in the self-complete socio-demographic questions at the end of the interview.

Respondents were interviewed in their own home by trained and experienced interviewers who had worked on previous valuation surveys, including the HUI2 (21), and OAB-5D (11). The project was approved by the ScHARR Research Ethics Committee at the University of Sheffield.

Analysis

Respondents are excluded from the analysis if they value all states as identical and less than one; value the worst possible health state higher than every other state; or value all states as worse than dead. Descriptive statistics of health state values across the three label groups are presented. Mean health state values are compared across the three condition labels using simple t-tests.

Regression techniques are used to estimate the impact of each label on health state values. The regression analysis further examines whether experience of the condition being valued impacts on health state values, whilst controlling for other socio-demographic characteristics. The standard specification is:

$$y_{ij} = \alpha + \boldsymbol{\beta}\mathbf{x}_j + \boldsymbol{\gamma}\mathbf{q}_i + \boldsymbol{\theta}\mathbf{r}_{ij} + \boldsymbol{\delta}\mathbf{z}_i + \varepsilon_{ij} \quad (1)$$

where $i = 1, 2, ..., n$ represents individual respondents and $j = 1, 2, ..., m$ represents the 8 health states. The dependent variable, *y*, represents the TTO utility value, **x**

represents the vector of dummies for the health states, **q** represents the vector of dummies to capture labelling effects, **r** represents the vector of interaction terms to capture labelling and severity effects, **z** represents the vector of socio-demographic characteristics including experience with the condition, and ε_{ij} represents the error term. All respondents have multiple observations and random effects generalized least squares (GLS) models are used to take account of this (22). Goodness of fit statistics are reported. STATA version 9 was used for all regression analysis and SPSS version 15 was used for the descriptive statistical analysis.

The data

Two hundred and forty one respondents from the North of England were successfully interviewed, providing a response rate of 39% of respondents who answered their door to the interviewer. The completion rate across all completed interviews was 99% for all TTO tasks. All responses were included in the analysis as no respondents met the exclusion criteria.

Characteristics of the overall sample are compared to the general population in South Yorkshire and England in Table 2. The study sample has a larger proportion of females, retired people and respondents aged over 65 and a lower proportion of self-employed and employed individuals. Mean EQ-5D score for the sample is lower than the general population in England (0.82 compared to 0.86 (23)).

Respondent characteristics vary significantly at the 5% level across the three groups regarding respondents aged 18-40 and full-time students (Table 3) using a factorial analysis of variance (ANOVA) estimated using a generalized linear model. The IBS label group has a relatively smaller proportion of respondents aged 18 to 40 and full time students and the no label group has a relatively higher proportion of full-time students. The cancer label group has a larger proportion of respondents with experience of the condition in their family and in caring for others than the IBS group, but a smaller proportion of respondents with experience of the condition in themselves. This may reflect the prevalence of the condition, but may also reflect that cancer is a more obvious condition to others.

Results

Descriptive statistics

Table 4 presents descriptive statistics of health state values across the three labelling groups: no label; IBS label; and cancer label. For 6 of the 8 health states the mean value is highest for the IBS label. For 6 of 8 states the mean value is lowest for the cancer label, and these are for the more severe health states. Mean, median and inter-quartile range for the two mildest states are similar across all 3 groups, yet differences are apparent for the cancer label group for more severe states and most noticeably for the worst (pits) state 54444444. In comparison to tariff values estimated in the original valuation study with no labels (20), mean health state values are lower for 7 of 8 states in the cancer label group, 6 of 8 states in the no label

group and 3 of 8 states in the IBS label group. However differences in mean health state value between the tariff value and values from the no label and IBS label groups are small, with the exception of the worst (pits) state 54444444. The proportion of observations that are worse than dead (below zero) varies from 5.0% in the no label group and 7.5% in the IBS label group to 12.7% in the cancer label group. The proportion of observations equal to 1 varies from 29.4% in the no label group and 27.0% in the IBS label group to 21.2% in the cancer label group.

Simple t tests reveal that health state values for the cancer label group are significantly different to health state values for both the no label group (P value = 0.006) and the IBS group (P value = 0.000) but that there are no significant differences between the no label and IBS groups (P value =0.283).

Regression analysis

Table 5 presents regression analysis examining the relationship between health state values, health state descriptions, labels and socio-demographics. Model (1) includes only state level dummy variables and label dummy variables; model (2) in addition incorporates socio-demographic variables; model (3) in addition includes experience of the labelled condition, model (4) adds to model (2) interaction terms to reflect the interaction between the specific health state and labelled condition; and model (5) combines variables included in models (3) and (4).

Regarding the socio-demographics and explanatory variables used in the regression analysis, spearman correlation coefficients indicate poor correlation (<|0.3| (24)) with the exception of moderate correlation (<|0.7| (24), here are <|0.5|) between employed and retired variables and between experience of IBS in family and in caring for others variables. This indicates that multicollinearity should not be a problem and that the use of these variables together is acceptable. Models incorporating a range of socio-demographic and experience variables were estimated and the best models (using diagnostics, correlations and proportion of significant coefficients) are presented here. The choice of variables was also informed using the results from ANOVA and related tests (not reported).

The results show that all state dummy variables are significant at the 1% level and are consistent, meaning that the utility decrement is larger for more severe health states as expected (where severity is calculated by summing the levels of all dimensions). Across all models the IBS label is never significant and the cancer label is significant in models (2) and (3) but not in model (1). Statistical measures of within, between and overall R-squared, root mean squared error, rho and Wald chi-squared indicate that models (2), (3), (4) and (5) perform better than model (1). This shows that the inclusion of socio-demographics variables and interaction terms rather than a simple additive labelling variable for cancer improves the model. The inclusion of experience of the labelled condition also improves the model. Models were estimated with interaction terms for IBS but these variables were never significant and did not improve the model.

Interaction terms for the state and the cancer label are significant for the five more severe states but not for the three least severe states. There is little difference in the coefficients for states 44321321, 23141224, 24432411 and 51224434, varying from a utility decrement of -0.093 to -0.114 in model (4). However the size of the utility decrement for the most severe (pits) state 54444444 is much larger at -0.219 in model (4). The inclusion of the interaction terms reduces the size of the coefficient for the state dummy variables with the exception of state 31212241. For this state, the interaction term is positive in model (4) but not significant. Variables representing experience of the labelled condition are significant for the cancer label but not for the IBS label. Experience of cancer in themselves has a utility decrement, whereas experience of cancer in caring for others has a significant utility increment in both models where it is included. Only two socio-demographic variables are significant; being a student reduces all utility values and being unemployed increases all utility values (although note that the number of respondents in each category is small).

Discussion

Our results demonstrate that the inclusion of a condition label can affect health state values, but this is dependent upon the specific condition. This is in accordance with the literature where some studies find that labelling affects health state values, whereas other studies find the reverse. Contrary to previous studies we find that the inclusion of a cancer label in health state descriptions affects health state values and we further find that the impact is dependent upon the severity of the state. When label and severity interactions are accounted for in the model, the inclusion of a cancer label does not significantly affect health state values for milder states, but has a significant reduction in health state values for more severe states (from 0.093 to 0.114 in model (4)) and a noticeably larger reduction for the worst state (0.219 in model (5)).

The inclusion of an IBS label rather than no condition label does not significantly affect health state values though the values are slightly raised, and this is consistent across all states of differing severity. The difference in impact on values across the two labels is striking. One possible reason causing the impact is that the prior knowledge and preconception of the condition is different in each case. IBS is a long-term chronic disorder affecting the digestive system that is not well publicised and generally regarded as mild. Cancer is used to refer to many different conditions affecting different parts of the body, and can be terminal. It is associated with fear and dread, and comes with preconceived ideas about prognosis that may influence values. This knowledge may affect the values assigned to health states, despite the fact that the quality of life and health state description is identical.

Another possibility is that the difference in condition represents a change in the underlying health state, and this is what affects values. For example, respondents may place a different value on interference with social activities due to, say, needing to be near a toilet than due to undergoing chemotherapy. This raises the question of whether health state values should reflect this difference. The aim of producing health state values is to inform resource allocation decisions across all conditions and patient groups, meaning there must be comparability across common health

state descriptions regardless of the underlying condition. Otherwise this could imply that, for example, a given generic EQ-5D state has a lower utility value for a patient with cancer than for a patient with IBS. We therefore argue that health state values developed for use in resource allocation should not reflect any difference in values introduced by referring to a medical condition. This poses difficulties for preference-based measures derived from existing condition-specific measures with items that mention the condition. Here a choice must be made between either maintaining the reference to the condition with the possibility of introducing a distortion in values, or removing the condition label meaning that the preference-based measure does not perfectly map to the original measure. Neither option is ideal. One alternative is to develop a descriptive system de novo, yet this is time consuming and expensive. Furthermore condition-specific preference-based measures are often developed from existing measures because of their wide usage and established reliability and validity, and sometimes because this enables utilities to be obtained from existing datasets where other options are unsuitable.

Each label group has a large proportion of respondents with experience of the condition (40% in the IBS group, 78% in the cancer group) and variables representing experience of cancer in themselves and in caring for others were significant in the regression analysis. This is problematic as it again suggests that the elicited values may not purely reflect the health state being valued. ANOVA tests (not reported) revealed that IBS values were affected by experience of IBS in the family or in caring for others, yet these variables were insignificant in the regression analysis. Experience of cancer in themselves leads to *lower* health state values, which is inconsistent with the literature examining patient values where it is found that patients provide higher health state values than the general population (25). Experience of cancer in caring for others leads to *higher* health state values, meaning greater unwillingness to sacrifice years of life in return for increased quality of life. This could be due to a variety of reasons such as knowledge of adaptation, although this would be incompatible with the previous finding.

Generic preference-based measures are claimed to be applicable for a wide range of patients and interventions, to allow comparability in economic evaluations undertaken across competing interventions. Yet generic measures are not always available or appropriate and in these circumstances condition-specific preference- based measures are one option that can be used to produce utility values for use in economic evaluation. To ensure comparability across economic evaluations health state values produced by different preference-based measures should reflect the health state description and should not be affected by the reference to the condition per se. We demonstrated that the condition label of cancer affected values whereas the IBS label did not, and hence further research is required to determine which condition labels may distort values. Until this information is available, it is recommended to avoid condition labels in health state descriptions to ensure that values are not distorted by prior knowledge or preconception of the condition. However for some conditions and existing measures this is impractical, as the condition label is embedded within the classification system.

Table 1: EORTC-8D classification system

During the past week:

Physical functioning
You had no trouble taking a long walk
You had a little trouble taking a long walk
You had quite a bit of trouble taking a long walk
You had very much trouble taking a long walk
You had very much trouble taking a short walk outside of the house

Role functioning
You were not limited in pursuing your hobbies or other leisure time activities
You were limited a little in pursuing your hobbies or other leisure time activities
You were limited quite a bit in pursuing your hobbies or other leisure time activities
You were limited very much in pursuing your hobbies or other leisure time activities

Social functioning
Your physical condition or medical treatment did not interfere with your social activities
Your physical condition or medical treatment interfered a little with your social activities
Your physical condition or medical treatment interfered quite a bit with your social activities
Your physical condition or medical treatment interfered very much with your social activities

Pain
Pain did not interfere with your daily activities
Pain interfered a little with your daily activities
Pain interfered quite a bit with your daily activities
Pain interfered very much with your daily activities

Emotional functioning
You did not feel depressed
You felt a little depressed
You felt quite a bit depressed
You felt depressed very much

Fatigue and sleep disturbance
You were not tired
You were a little tired
You were quite a bit tired
You were tired very much

Constipation and diarrhoea
You were not constipated and did not have diarrhoea
You were constipated and/or had diarrhoea a little
You were constipated and/or had diarrhoea quite a bit
You were constipated and/or had diarrhoea very much

Nausea
You did not feel nauseated
You felt a little nauseated
You felt nauseated quite a bit
You felt nauseated very much

Table 2: Respondent characteristics

	Sample (n=241)	*South Yorkshire*[1]	*England*[2]
Mean age (s.d.)	49.95 (19.84)	NA	NA
Age distribution			
18-40	39.8%	41.2%	41.6%
41-65	32.8%	39.1%	39.1%
Over 65	25.3%	19.7%	19.3%
Female	60.2%	51.2%	51.3%
Married/Partner	57.3%	NA	NA
Employed or self-employed	38.6%	56.1%	60.9%
Unemployed	3.7%	4.1%	3.4%
Long-term sick	5.8%	7.7%	5.3%
Full-time student	7.1%	7.5%	7.3%
Retired	28.2%	14.4%	13.5%
Own home outright or with a mortgage	69.3%	64.0%	68.7%
Renting property	30.3%	36.0%	31.3%
Secondary school is highest level of education	44.0%	NA	NA
EQ-5D score (s.d.)	0.82 (0.25)	NA	0.86 (0.23)[2]

[1] Statistics for South Yorkshire Health Authority and 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.

[2] Interviews conducted in the Measurement and Valuation of Health (MVH) study in 1993 (23).

Table 3: Respondent characteristics by label

	No label (n=81)	*IBS label (n=80)*	*Cancer label (n=80)*	*ANOVA P-value*
Mean age (s.d.)	48.08 (21.57)	51.78 (18.93)	47.1 (18.85)	0.300
Age distribution				
18-40	45.6%	28.8%	46.3%	0.045
41-65	26.6%	40.0%	32.5%	0.165
Over 65	27.8%	27.5%	21.3%	0.316
Female	61.7%	62.5%	56.3%	0.681
Married/partner	63.0%	52.5%	56.3%	0.400
Main activity				
Employed or self-employed	35.8%	38.8%	41.3%	0.779
Unemployed	2.5%	2.5%	6.3%	0.350
Long-term sick	7.4%	3.8%	6.3%	0.602
Full-time student	12.3%	2.5%	6.3%	0.048
Housework	8.6%	13.8%	15.0%	0.435
Retired	29.6%	32.5%	22.5%	0.354
Own home outright or with a mortgage	71.3%	67.5%	70.0%	0.913
Renting property	28.7%	32.5%	30.0%	0.851
Secondary school is highest level of education	48.1%	43.8%	40.0%	0.584
EQ-5D score (s.d.)	0.78 (0.30)	0.80 (0.27)	0.85 (0.20)	0.290
TTO completion rate	98.6%	99.4%	99.5%	
Doubtful whether respondent understood TTO (interviewer reported)	3.8%	3.9%	1.3%	
Experience of labelled condition		40.0%	78.0%	
in themselves		18.8%	12.5%	
in family		27.5%	67.5%	
in caring for others		13.8%	36.3%	
Time taken	32.7 (10.3)	32.8 (9.3)	33.8 (9.5)	

Table 4: Descriptive statistics of health state values across all labelling groups

	Original study (20) *(n=344)*	*No label (n=81)*			*IBS label (n=79-80)*[1]			*Cancer label (n=79-80)*[1]		
Health state	*Tariff value*	*Mean (s.d.)*	*Median*	*IQR*	*Mean (s.d.)*	*Median*	*IQR*	*Mean (s.d.)*	*Median*	*IQR*
11111111	1	0.96 (0.13)	1.00	1.00 - 1.00	0.99 (0.06)	1.00	1.00 - 1.00	0.96 (0.12)	1.00	1.00 - 1.00
31212241	0.75	0.74 (0.32)	0.83	0.61 - 1.00	0.81 (0.23)	0.88	0.70 - 1.00	0.80 (0.22)	0.88	0.68 - 1.00
13423411	0.72	0.67 (0.30)	0.73	0.49 - 1.00	0.71 (0.37)	0.83	0.60 - 0.93	0.64 (0.36)	0.78	0.46 - 0.93
44321321	0.65	0.66 (0.35)	0.73	0.48 - 1.00	0.68 (0.37)	0.76	0.58 - 0.93	0.56 (0.50)	0.70	0.48 - 0.93
23141224	0.64	0.63 (0.36)	0.69	0.48 - 1.00	0.69 (0.36)	0.79	0.53 - 0.98	0.57 (0.45)	0.68	0.48 - 0.80
24432411	0.64	0.66 (0.33)	0.73	0.46 - 1.00	0.65 (0.40)	0.78	0.50 - 0.94	0.54 (0.44)	0.66	0.38 - 0.83
51224434	0.51	0.49 (0.41)	0.53	0.26 - 0.84	0.53 (0.42)	0.60	0.38 - 0.83	0.41 (0.49)	0.50	0.23 - 0.78
54444444	0.29	0.20 (0.49)	0.18	0.00 - 0.54	0.17 (0.49)	0.16	-0.03 - 0.50	-0.03 (0.50)	0.00	-0.38 - 0.38

[1] Eighty observations for all states with the exception of states 11111111 and 13423411 for the IBS label group and states 11111111, 31212241 and 54444444 for the cancer label group.

Table 5: Regression analysis of health state values across different labelling groups

	(1)	(2)	(3)	(4)	(5)
States					
31212241	-0.187***	-0.187***	-0.187***	-0.197***	-0.197***
13423411	-0.297***	-0.297***	-0.297***	-0.284***	-0.284***
44321321	-0.340***	-0.340***	-0.340***	-0.304***	-0.303***
23141224	-0.343***	-0.343***	-0.343***	-0.313***	-0.313***
24432411	-0.354***	-0.354***	-0.353***	-0.317***	-0.317***
51224434	-0.489***	-0.489***	-0.489***	-0.456***	-0.456***
54444444	-0.856***	-0.856***	-0.856***	-0.785***	-0.784***
Labelling					
IBS	0.027	0.015	-0.009	0.014	-0.009
Cancer	-0.067	-0.083**	-0.115**		
Cancer interaction terms					
11111111 x Cancer				-0.006	-0.038
31212241 x Cancer				0.027	-0.005
13423411 x Cancer				-0.044	-0.076
44321321 x Cancer				-0.112**	-0.145**
23141224 x Cancer				-0.093*	-0.126**
24432411 x Cancer				-0.114**	-0.146**
51224434 x Cancer				-0.102*	-0.134**
54444444 x Cancer				-0.219***	-0.251***
Experience of labelled condition					
Cancer in themselves			-0.149		-0.150*
Cancer in caring for others			0.142**		0.142**
IBS in caring for others			0.037		0.036
IBS in family			0.072		0.073
Socio-demographics					
Female		0.011	0.007	0.010	0.007
Married		0.055	0.056	0.055	0.056
Retired		-0.036	-0.024	-0.036	-0.024

	(1)	(2)	(3)	(4)	(5)
Unemployed		0.119	0.135*	0.119	0.135*
Student		-0.244***	-0.241***	-0.245***	-0.241***
Housework		0.003	0.008	0.002	0.008
Long term sick		0.063	0.098	0.063	0.098
Constant	0.983***	0.972***	0.966***	0.946***	0.940***
Observations	1910	1910	1910	1910	1910
Number of id	241	241	241	241	241
Within R-squared	0.453	0.453	0.453	0.462	0.462
Between R-squared	0.028	0.115	0.149	0.115	0.149
Overall R-squared	0.028	0.115	0.149	0.115	0.149
Root MSE	0.271	0.271	0.272	0.270	0.270
Wald Chi-squared	1376.873	1397.769	1406.900	1445.517	1454.406

Notes: *** significant at 1% level, ** significant at 5% level, * significant at 10% level.

Reference state is 11111111 valued with no label.

Appendix: Information sheets about IBS and cancer

What is irritable bowel syndrome?

Irritable bowel syndrome (IBS) is a chronic (long-term) disorder that affects the digestive system. It causes abdominal pain, diarrhoea and constipation. There are different types of irritable bowel syndrome, depending on your main symptom.

Taken from the NHS Direct website, see www.nhs.uk

What is cancer?

Cancer is a term that is used to refer to a number of conditions where the body's cells begin to grow and reproduce in an uncontrollable way. This rapid growth of cancerous cells is known as a malignant tumour. These cells can then invade and destroy healthy tissue, including organs. There are hundreds of different types of cancer.

Taken from the NHS Direct website, see www.nhs.uk

References

(1) Dolan P. Modeling valuations for EuroQol health states. Medical care 1997 Nov;35(11):1095-108.

(2) National Institute of Health and Clinical Excellence (NICE). Guide to the methods of technology appraisal. London: NICE; 2008.

(3) Marra CA, Woolcott JC, Kopec JA, Shojania K, Offer R, Brazier JE, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. Soc Sci Med 2005 Apr;60(7):1571-82.

(4) Espallargues M, Czoski-Murray CJ, Bansback NJ, Carlton J, Lewis GM, Hughes LA, et al. The impact of age-related macular degeneration on health status utility values. Investigative Ophthalmology & Visual Science 2005 Nov;46(11):4016-23.

(5) Barton GR, Bankart J, Davis AC, Summerfield QA. Comparing Utility Scores Before and After Hearing-Aid Provision : Results According to the EQ-5D, HUI3 and SF-6D. Applied Health Economics & Health Policy 2004;3(2):103-5.

(6) Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. European Journal of Health Economics 2010 Apr;11(2):215-25.

(7) Bennett KJT, Torrance GW, Boyle M, Guscott R, Moran L. Development and testing of a utility measure for major, unipolar depression (McSad). Quality of Life Research 2000;9(1):109-20.

(8) Bennett KJT, Torrance GW, Moran L, Smith F, Goldsmith CH. Health state utilities in knee replacement surgery: The development and evaluation of McKnee. Journal of Rheumatology 1997;24(9):1796-805.

(9) Yang Y, Tsuchiya A, Brazier J, Young T. Estimating a preference-based single index from the Asthma Quality of Life Questionnaire (AQLQ). Health Economics and Decision Sciences Discussion Paper 07/02, University of Sheffield 2007.

(10) Young T, Yang Y, Brazier J, Tsuchiya A. The use of Rasch analysis as a tool in the construction of a preference based measure: the case of AQLQ. Health Economics and Decision Science Discussion Paper 07/01 ScHARR, University of Sheffield 2007.

(11) Yang Y, Brazier J, Tsuchiya A, Coyne K. Estimating a preference-based single index from the overactive bladder questionnaire. Value in Health 2009;12(1):159-66.

(12) Young T, Yang Y, Brazier JE, Tsuchiya A, Coyne K, Young T, et al. The first stage of developing preference-based measures: constructing a health-state classification using Rasch analysis. Quality of Life Research 2009 Mar;18(2):253-65.

(13) Brazier J, Tsuchiya A. Preference-based condition-specific measures of health: what happens to cross programme comparability? Health Econ 2010 Feb;19(2):125-9.

(14) Brazier J, Rowen D, Tsuchiya A, Yang Y, Young T. What a pain: adding a generic dimension to a condition-specific preference-based measure. Paper presented at HESG meeting, University College Cork, June 2010.

(15) Gerard K, Dobson M, Hall J. Framing and labelling effects in health descriptions: quality adjusted life years for treatment of breast cancer. Journal of clinical epidemiology 1993 Jan;46(1):77-84.

(16) Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, Ciampi A, Till JE, Boyd NF. Describing health states. Methodologic issues in obtaining values for health states. Medical care 1984 Jun;22(6):543-52.

(17) Rabin R, Rosser RM, Butler C. Impact of diagnosis on utilities assigned to states of illness. Journal of the Royal Society of Medicine 1993 Aug;86(8):444-8.

(18) Robinson S, Bryan S. 'Naming and framing': an investigation of the effect of disease labels on health state valuations. Health Economics Study Group Meeting, University of Oxford 2001.

(19) Sackett DL, Torrance GW. The utility of different health states as perceived by the general public. Journal of Chronic Diseases 1978;31(11):697-704.

(20) Rowen D, Brazier J, Young T, Gaugris S, Craig BM, King MT, et al. Deriving a preference-based measure for cancer using the EORTC QLQ-C30. Health Economics and Decision Science Discussion Paper, University of Sheffield 2010.

(21) McCabe C, Stevens K, Roberts J, Brazier J. Health state values for the HUI 2 descriptive system: results from a UK survey. Health Econ 2005 Mar;14(3):231-44.

(22) Brazier J, Roberts J, Deverill M, Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics 2002 Mar;21(2):271-92.

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

(24) Tabachnick BG, Fidell LS. Using Multivariate Statistics. Boston: Allyn and Bacon; 2001.

(25) Dolan P, Kahneman D. Interpretations of utilities and their implications for the valuation of health. Economic Journal 2008;118:215-34.