
Article:
Rowen, D., Brazier, J. and Van Hout, B. (2014) A Comparison of Methods for Converting DCE Values onto the Full Health-Dead QALY Scale. Medical Decision Making. Published Online, November 14, 2014.

https://doi.org/10.1177/0272989X14559542

A comparison of methods for converting DCE values onto the full health-dead QALY scale

Donna Rowen[a], John Brazier[a], Ben Van Hout[a]

[a] School of Health and Related Research (ScHARR), University of Sheffield
* 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

Key words: Preference-based measures of health; Health-related quality of life; Discrete Choice Experiment; Pairwise comparisons

Funding source: Financial support for the data collection of data analysed in this study was provided by Novartis. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing this report.

ABSTRACT

Background

Preference elicitation techniques such as Time Trade-Off (TTO) and Standard Gamble (SG) receive criticism for their complexity and difficulties in use. Ordinal techniques such as Discrete Choice Experiment (DCE) are arguably easier to understand, but generate values that are not anchored onto the full health-dead 1-0 QALY scale required for use in economic evaluation.

Methods

This paper compares existing methods for converting modelled DCE latent values onto the full health-dead QALY scale: (1) anchoring DCE values using dead as valued in the DCE; (2) anchoring DCE values using TTO value for worst state; to two new methods: (3) mapping DCE values onto TTO; (4) combining DCE and TTO data in a hybrid model. Models are compared using their ability to predict mean TTO health state values.

Data

We use postal DCE data (n=263) and TTO data (n=307) collected by interview in a general population valuation study of an asthma condition-specific measure (AQL-5D).

Results

New methods (3) and (4) using mapping and hybrid models are better able to predict mean TTO health state values (mean absolute difference (MAD) 0.052 to 0.084) than the anchor-based methods (MAD 0.075 to 0.093) and were better able to predict mean TTO health state values even when using in their estimation a subsample of the available TTO data.

Conclusions

These new mapping and hybrid methods have a potentially useful role for producing values on the QALY scale from data elicited using ordinal techniques such as DCE for use in economic evaluation that makes best use of the desirable properties of each elicitation technique and elicited data. Further research is encouraged.

1. INTRODUCTION

Economic evaluation measuring outcomes using quality-adjusted life years (QALYs) has increasingly informed resource allocation in recent years. The QALY is a measure of health outcome that combines quality of life with length of life. Quality of life is measured on a full health-dead 1-0 scale, where one equals full health and zero is equal to being dead, with negative values where quality of life is regarded as worse than being dead. The QALY enables comparisons across interventions that impact on mortality, morbidity and both. These comparisons cannot be avoided, and the QALY provides a useful summary measure that enables cost effectiveness analysis. However there remains much debate surrounding the elicitation of utilities to produce the 'Q' quality adjustment weight of the QALY.

Standard cardinal techniques for eliciting preferences for health states are time trade-off (TTO), standard gamble (SG) and visual analogue scale (VAS). There has been much debate in the literature regarding the best technique. TTO and SG have been regarded by many as superior to VAS for eliciting preferences since they are based on choices that involve sacrificing (i.e. requires the respondents to express a willingness to trade). Given certain assumptions hold, then TTO and SG have been shown to elicit values from an underlying utility function (1, 2). However, SG and TTO have been criticised for being too complex for many respondents. This has led to increasing interest in the use of ordinal techniques, such as pairwise discrete choice experiment (DCE) where respondents choose between two health states and best-worst scaling (BWS) where respondents choose the best and worst attributes of a health state, including application of BWS to measures of capabilities (3) and social care outcomes (4).

DCE has recently gained popularity for eliciting health state utility values to inform the scoring systems for preference-based measures of health. A small number of studies have applied DCE to value health states (5-10) but the majority of these have not anchored values

on the full health-dead QALY scale. DCE values can be modelled using regression analysis to produce preference weights for each severity level of each dimension in the classification system, but these coefficients are expressed on a modelled latent utility value that has arbitrary anchors. Some studies anchor values onto a 1-0 best state-worst state scale (5, 9) but this is dependent on the specific dimensions and severity levels included in the classification system and is not on the QALY scale. For health care programmes where survival and quality of life are important outcomes this method will not be sufficient as values must be on the full-health dead QALY scale to generate a single outcome measure for cost-effectiveness assessments.

Existing DCE studies have anchored modelled latent utility values onto the full-health QALY scale using three methods. The first method uses the TTO value for the worst health state defined by the classification system (10). The second method includes a dead state into pairwise DCE tasks and estimates an additive regression model that includes a dummy variable for dead (7)). The regression coefficients are normalised onto the full health-dead scale by dividing the coefficient on each level by the coefficient on the dead dummy variable. This method has been criticised (11) as many respondents do not see any of the states described by the classification system as worse than being dead. While the proportion who do not regard any state as worse than dead was only 15% in the UK valuation of the EQ-5D, it can be more than half (e.g. 66% for SF-6D) and for these respondents the appropriateness of this method is more questionable. The third method incorporates duration into the DCE task (12, 13). The second and third methods have the advantage that they enable the DCE data to be anchored onto the 1-0 full health-dead scale with no further information required. However in the third method the addition of duration as one of the attributes makes the DCE task more complex and means an even larger amount of information is involved in a single pairwise task. This third method is not included in this study.

The aim of our paper is to compare existing methods for anchoring DCE values onto the full health-dead QALY scale to two new methods proposed here. It uses a data set from a general population valuation study of an asthma-specific measure (AQL-5D) using TTO and DCE. The alternative methods being compared are existing methods: (1) anchoring the modelled DCE latent variable using dead as valued in the DCE; (2) anchoring the modelled DCE latent variable using the TTO value for the worst state; and new methods (3) mapping the modelled DCE latent variable onto TTO; (4) combining DCE and a sample of TTO data in a hybrid model. It is hypothesised that the two new proposed approaches that make better use of the DCE and TTO elicitation techniques and data will provide promising alternatives to the existing methods used in the literature.

2. METHODS

2.1 Health state description

Health states are described using an asthma-specific preference-based measure, AQL-5D, (14) derived from the Asthma Quality of Life Questionnaire, AQLQ (15). AQL-5D has 5 dimensions: concern about asthma, shortness of breath, weather and pollution stimuli, sleep impact and activity limitations. Each dimension has 5 levels of severity to define a total of 3125 health states (see Table 1). Health states consist of one severity level per dimension, for example the worst state, state 55555, is made up of the most severe level of each dimension.

2.2 Valuation surveys

2.2.1 Interview

Interviews were undertaken to elicit health state utility values for a selection of AQL-5D states using TTO from a representative sample of the general population. Respondents were interviewed in their own home by trained and experienced interviewers. At the start of the interview respondents were informed about asthma using an information sheet. To

familiarise respondents with the system respondents who had asthma were asked to complete the health state classification system for themselves; respondents who did not have asthma were asked to complete the health state classification system for someone they knew who had asthma or to imagine somebody with asthma. Respondents were then asked to rank 7 intermediate states, full health (health state 11111), worst state defined by the health state classification (health state 55555), and immediate death.

Respondents then valued a practice state using TTO, and this was followed by valuation of the 7 intermediate states and the worst state using TTO, with an upper anchor of measure-specific full health (health state 11111). The survey used the Measurement and Valuation of Health (MVH) study version of TTO, including a visual prop designed by the MVH group (University of York) (16, 17) (see Appendix 1) which determines the point where respondents are indifferent between 10 years in the impaired health state and *x* years in state 11111. For states regarded as worse than dead indifference is determined between death and *y* years in the impaired health state followed by *x* number of years in state 11111 (where $x+y=10$). Respondents were then asked questions about their socio-economic characteristics and health service use, how difficult they found the rank and TTO tasks and finally whether they were willing to participate in a postal survey (described below). Respondents were not allowed to change their answers to any questions.

The classification system describes too many states for valuation. A sample of states were selected for valuation to enable the specification of an additive regression model estimating preference weights for all severity levels of each dimension in the classification system, using level 1 as the baseline. Health states were selected using a balanced design, which ensured that every level of every dimension had an equal chance of being combined with each severity level of the other dimensions. The design selected 98 health states which were then randomly stratified into mixed severity groups of 7 states based on their summed

severity score (summing the scores on all 5 dimensions e.g. 22222 has a severity score of 10). These combinations of 7 states made up the card blocks used in the interviews, to ensure each intermediate state was valued an equal number of times and that respondents valued states with a wide range of severity. The worst state was valued by all respondents to increase accuracy for this value and enable responses to be compared across groups of respondents valuing different intermediate states. The ordering of health states in the rank and TTO tasks was random for each respondent. The sample size was selected to ensure that each health state was valued a minimum of 20 times.

2.2.2 Postal survey

Interviewees who had stated in the interview that they were willing to complete a postal survey were mailed a postal DCE questionnaire approximately four weeks after the interviews. The questionnaire was also mailed to a sample of the general public who had not been previously interviewed. At the start of the survey respondents were asked to complete the health state classification system for themselves to help familiarise them with the classification. Respondents were then asked a practice pairwise comparison question followed by a series of 8 pairwise comparisons, where for each comparison respondents were asked to indicate which health state they preferred. An example is provided in Figure 1. Finally respondents were asked questions about their socio-demographic characteristics. Reminders were sent to all non-responders approximately four weeks after the initial questionnaire was sent.

Combinations of health states for the pairwise comparisons were selected using the D-efficiency approach using a specially developed programme (18) in statistical software SAS. The programme selected 24 pairwise comparisons which were randomly allocated to four questionnaire versions each with 6 comparisons. Each questionnaire also included two identical pairwise comparisons comparing a severe health state (state 44355) and the worst

health state to 'immediate death'. The number of questionnaires mailed out was determined using a targeted sample size ensuring that each pair of profiles was valued a minimum of 20 times (19) and expected response rate alongside funding constraints.

2.3 Modelling health state values

2.3.1 Time trade-off

TTO data was analysed using a one-way error components random effects model via generalised least squares (GLS). This takes account of the structure of the data as each respondent valued multiple health states (20). The model specification was:

$$U_{ij} = f(\boldsymbol{\beta}\mathbf{X}_{ij}) + \varepsilon_{ij} \quad (1)$$

Where U_{ij} represents TTO disvalue (1–TTO value) for health state j=1,2 …n valued by respondent i=1,2…m, $\mathbf{X}_i$ represents a vector of dummy variables for each level λ of dimension ∂ of the health state classification system where level λ = 1 is the baseline for each dimension and ε_{ij} is the error term. The error term is subdivided $\varepsilon_{ij} = u_i + e_{ij}$, where u_i is the individual random effect and e_{ij} is the usual random error term for the jth health state valuation of the ith individual. Other models estimated using this data are reported elsewhere (21).

2.3.2 DCE

DCE data was modelled to produce cardinal utility estimates on a latent utility scale. The DCE data was modelled using effects coding using a random effects probit model which takes account of the structure of the data where each respondent valued multiple states, using an additive specification as outlined in equation (1) (10). This model produced

coefficients on a latent utility scale with arbitrary anchors. This model excluded data collected for the pairwise comparisons involving the state 'dead'.

2.4 Translating DCE scores onto the full health-dead scale

2.4.1 Method (1): anchoring using the coefficient for 'dead'

Firstly all DCE data including data for the pairwise comparisons involving the state 'dead' was modelled using a random effects probit model (7). The model specification was:

$$U_{ij} = f(\boldsymbol{\beta}\mathbf{X}_{ij} + \phi D) + \varepsilon_{ij} \quad (2)$$

Where U_{ij} represents utility for health state *j*=1,2 …*n* valued by respondent *i*=1,2…*m*, $\mathbf{X}_i$ represents a vector of dummy variables for each level *λ* of dimension *∂* of the health state classification system, *D* represents a dummy variable for the state 'dead' and ε_{ij} is the error term. Secondly coefficients for each level of each dimension were normalised by dividing by the coefficient of the dead dummy variable; $\beta_{r\lambda\partial} = \beta_{\lambda\partial} / \Phi$ where $\beta_{r\lambda\partial}$ is the rescaled coefficient for level *λ* of dimension *∂*, $\beta_{\lambda\partial}$ is the coefficient for level *λ* of dimension *∂* and Φ is the coefficient for the dead dummy variable (see (7) for use of this technique in DCE data and (22, 23) for use of this technique in rank data).

2.4.2 Method (2): anchoring the worst state using TTO

This means that the value of the worst state in the DCE model is anchored at the TTO value of the worst state. The coefficients on a latent utility scale estimated in the random effects probit model of the DCE data were normalised onto the full health-dead scale using the estimated TTO value of the worst state. This is achieved using $\beta_{r\lambda\partial} = \beta_{\lambda\partial} \cdot w_{TTO} / w_{DCE}$, where

$\beta_{r\lambda\partial}$ is the rescaled coefficient for level λ of dimension ∂, $\beta_{\lambda\partial}$ is the coefficient for level λ of dimension ∂, w_{TTO} is the estimated TTO value for the worst state generated using the TTO model specified in equation (1) and w_{DCE} is the DCE value for the worst state estimated using the DCE model specified in equation (1).

2.4.3 Method (3): mapping DCE onto TTO

Mapping is a method often used to estimate utility values for a clinical trial (or other study of interest) when the utility measure of interest was not included. This is achieved by predicting utility values for the clinical trial using the statistical relationship between data that was included in the trial and the preferred utility measure estimated in an external dataset (see (24) for a recent review of mapping). This mapping approach is used here to estimate TTO values for all states using modelled latent DCE values for all states.

The probit model estimated on DCE data generates values on a latent utility scale for all 3125 states. Ninety-nine of these states also have mean TTO values collected from the interviews. The simple mapping function from TTO to DCE was specified as:

$$TTO_j = f(DCE_j) + \varepsilon_j \quad (3)$$

Where TTO_j represents mean TTO value of health state *j*, DCE represents the modelled latent utility value for health state *j* and ε_j is the error term. The first specification assumes a linear relationship with an intercept, and then squared and cubic terms were added to see whether model performance was improved. Estimation was undertaken using OLS on the mean health state values.

One purpose of this mapping might be to predict TTO values from DCE based on a mapping function estimated on TTO values for a small sub-set of states. For practical reasons the smaller the sub-set the better as this reduces the sample size of the TTO data which can be

expensive and time-consuming to collect. To explore the impact of sub-set size we examined mapping functions estimated from 10, 19 and 99 states. Method (3a) used 10 health states selected by ordering latent DCE utility estimates by severity (using the modelled DCE latent estimate) from mildest to most severe and selecting the states valued 1st, 11th, 22nd, 33rd, 44th, 55th, 66th, 77th, 88th and 99th. Method (3b) used 19 states, supplementing the states used in method 3a by states valued 6th, 16th, 27th, 38th, 49th, 60th, 71st, 82nd and 93rd. The rationale for choosing 10 and 19 states was logistical; these states can be easily valued by respondents using TTO. The study design for method (3a) requires respondents to value all 10 states using TTO; study design for method (3b) requires respondents to value 10 states, consisting of 9 states plus the worst state using two different card blocs in the interviews. Method (3c) used all 99 states in order to examine the degree of improvement from increasing the number of states valued by TTO up to the number required to estimate a well specified TTO algorithm.

2.4.4 Method (4): hybrid models

The generalised linear regression on the TTO data and the Probit regression on the DCE data contain a similar linear component $\mathbf{x}'\boldsymbol{\beta}$ underlying the TTO values and pairwise choices. This method assumes that this component, which reflects the weight given to the dimensions and labels, is the same between both models. This approach estimates the parameters by creating just one likelihood function which is the product of the likelihood of the TTO data and the likelihood of the DCE data. When optimizing this joint distribution one additional parameter is included that allows both models to differ up to a monotonic transformation and allows the variability to be different. This is done by a single parameter relating both linear functions with each other and by assuming different variances for the heterogeneity (or errors) in the TTO data and the DCE data. The hybrid method has similarities to information integration theory which "refers to the process whereby the psychological values of several stimuli are combined to produce a single impression". (25)

Here there are two elicitation methods that use different techniques to elicit utility values that represent the underlying preferences of each individual. The elicited TTO and DCE values are collectively modelled in a hybrid model to jointly inform the estimation of the modelled utility value, where TTO values provide cardinal information and DCE values provide ordinal information.

We present a random effects approach and a fixed effects approach to estimation. The likelihood of the fixed effects model is much easier to program (in R) and optimise than that of the random effects equivalent. For this reason a Bayesian method is used for the latter using WinBugs to obtain estimates of the parameters for the random effects case. Within the Bayesian model we choose non-informative priors which will generally lead to the same parameters estimates as a non-Bayesian approach. Methods (4a), (4b) and (4c) use individual level TTO data for the states selected in method (3) and all DCE data. Further technical details are presented in the Appendix 2.

2.5 Comparison of models

There is no gold standard with which to compare the performance of these models. However, where the aim is to predict TTO values then there will be interest in the degree of agreement between the values predicted by these 4 methods and observed TTO values. Agreement was assessed using mean absolute difference between observed TTO and predicted health state utility values (MAD) at the health state level, root mean squared difference (RMSD) at the health state level and number of states where MAD is greater than 0.05 and 0.1 respectively. Predictions from the 4 methods (and their variations) were plotted alongside mean observed values for the 99 states valued by TTO.

3. RESULTS

3.1 The data

The TTO dataset contains 307 successfully conducted interviews, providing a response rate of 40% for suitable respondents answering their door at time of interview. Each intermediate health state was valued 19 to 22 times, and the worst state was valued 307 times. The distribution of TTO values was negatively skewed and mean TTO values for the 99 health states ranged from 0.39 to 0.94. Further details are reported elsewhere (21).

The DCE dataset contains 263 successfully completed postal surveys. Out of the 307 respondents who were interviewed 168 returned postal DCE questionnaires (55%). Out of the 400 households receiving the postal questionnaire who were not previously interviewed 95 returned questionnaires (24% return rate). Data from all respondents have been pooled since previous analyses showed no significant difference between them (7).

Table 2 reports the socio-demographic composition of the TTO and DCE samples. Both samples are similar, but the TTO sample is younger and healthier, with a higher proportion of males. Self-reported health status using EQ-5D (16) was similar for each sample to the UK EQ-5D norms of 0.85 for females and 0.86 for males (26). The age distribution is significantly different across the two samples. No respondents were excluded from the analyses.

3.2 TTO model

Table 3 reports the model estimated on TTO data. The majority of coefficients were negative and the size of coefficients monotonically increased with more severe levels of each dimension. Three coefficients were positive but small and statistically insignificant.

3.3 DCE model

The DCE model producing latent DCE estimates that are unanchored onto a full health-dead 1-0 scale is reported in Table 3. Estimated coefficients for both methods had four

inconsistencies, level 2 of concern, breathlessness and pollution and level 3 of pollution, though only level 2 of pollution was statistically significant.

3.4.1 Methods (1) to (2): anchoring

Method (2) anchors the latent DCE estimates, and the results have the same non-monotonic coefficients as the latent estimates for level 2 of concern, breathlessness and pollution and level 3 of pollution (Table 3). For the 2 anchor based methods the pattern of coefficients was similar and mostly consistent with the levels of the dimensions (apart from the pollution dimension). The most noticeable differences were at the lower end of the dimensions for concern, short of breath, sleep and activities where method (1) produced larger coefficients than method (2) and the TTO model.

3.4.2 Method 3: mapping

The DCE coefficient is negative and significant across all 3 models (Table 4). The size of the intercept and the gradient associated with the latent DCE value are similar across models using TTO data collected for 10, 19 and 99 health states (models (3a), (3b) and (3c) respectively). Plots of TTO and DCE data indicated a linear relationship (7). The inclusion of squared and cubic terms was explored but these variables were insignificant and did not improve model performance.

3.4.3 Method 4: hybrid models

Results for method (4) are reported in Table 5 using both a common likelihood function and a Bayesian method. Overall the coefficients are similar to the TTO model reported in Table 3. Coefficients were larger for sleep and activity level 5 than in the TTO model, as also found for the anchor based models. There was a tendency for the coefficients to move in the direction of the anchor based models with larger coefficients for concern, sleep and activity levels 5, but this was less marked and was not the case for breathing. This tendency was

greater for the two models with sub-samples of TTO valued states. For the likelihood model estimates using TTO data for 10 and 19 states alongside all DCE data there are 3 inconsistencies, though none are significant. The comparable models estimated using the Bayesian method have 4 and 5 inconsistent coefficients and these are significant for level 2 weather and pollution using method 4 and level 2 short of breath and level 2 sleep using method 4c. The inconsistent coefficient for level 2 of the weather and pollution dimension was also found in the models estimated using methods 1 and 2 but is only significant in model 4. This suggests that respondents may not have a difference in preference for levels 1 and 2 for the weather and pollution dimension meaning that it is acceptable to merge level 2 with the reference level for this dimension.

3.5 Comparison of methods

The smallest difference between predicted and observed mean TTO health state values measured using MAD and RMSD was methods (3c) and (4c). However simple mapping functions using 10 and 19 mean TTO health state values produced similar levels of agreement. Hybrid models estimated using TTO values for 10 and 19 states produced larger differences with models using the likelihood method than the Bayesian method. Method (2) was the closest of the anchor models.

These differences can be seen in Figure 2. Method (3) produced the utility estimates that best follow most closely the pattern of observed TTO values. Method (1) under-estimates at the lower end of the scale. Method (2) over-estimated the value of most states. Method (4) over-estimated values for the majority of health states, but to a less severe extent than method (2).

4. DISCUSSION

This paper explored two new methods for converting modelled DCE latent values for a health state classification system onto the full health-dead 1-0 QALY scale and compared these to existing methods from the literature. The first new method mapped modelled DCE latent values onto TTO values. The second method estimated utility decrements for each severity level of the classification system by modelling DCE and TTO data together using a hybrid model. These new methods produce utility estimates that are more similar to TTO estimates than existing methods, and are arguably more appropriate for anchoring DCE values onto the full health-dead QALY scale. The analysis further explored whether these methods would produce accurate utility estimates for studies involving a small-scale TTO survey alongside a large DCE survey. Both of these new methods produced relatively similar predictions to a larger scale TTO survey under these circumstances. These results have not been compared to the results for DCE with duration as an additional attribute, although future research comparing the mapping and hybrid approaches to DCE with duration is encouraged.

These new methods potentially have a useful role in producing values on the QALY scale from DCE and TTO data that make best use of the desirable properties of each elicitation technique and elicited data. DCE has the advantage that the task itself is arguably easy to understand and values are not affected by time preference; but requires respondents to simultaneously consider a large amount of information and faces the challenge of how to convert values onto the full health-dead scale. TTO encapsulates the trade-off between quality and quantity of life; but is a complex task and data can be expensive and time-consuming to collect. Combining these techniques enables large scale data collection using DCE to be undertaken inexpensively online with small scale TTO data collected by interview as its usability in an online environment is questionable. There has also recently been interest in using DCE to obtain values from children (27) and elderly users of social care (28) and these new methods offer a way of using this data to produce values on the QALY scale.

Anchoring methods (1) and (2) used in the literature did not predict TTO as well as the new methods. Method (1) involves the use of pairwise comparisons with the state 'immediate death' and is an adaptation of a method successfully applied in rank data for several generic preference-based measures. Using SF-6D and HUI2 rank data the equivalent regression model was able to predict mean SG health state values reasonably well (22). However when using EQ-5D rank data the same model over-predicted TTO health state values (23), and model (1) estimated here replicated these results. The model has also been criticised since it violates the assumptions of random utility theory for the large proportion of respondents who do not value any state as worse than being dead (11). However it may have a role in more severe health state descriptive systems such as EQ-5D. Method (2) anchored the DCE data using a single data point for the TTO value of worst state. This method systematically over-estimated values due to its reliance on a single TTO value.

Method (3) used a simple mapping based approach and achieved good predictions of observed mean TTO health states values. This was largely unaffected when the method was estimated using datasets containing TTO values for only 10 and 19 health states respectively. Perhaps this is not surprising since it uses mean values and so removes much of the underlying variability. However there is considerable uncertainty around these mapped mean health state values. This would need further investigation before these results can be used in economic evaluation, for example using bootstrapping methods to generate confidence limits around these results.

Method (4) used a hybrid model to combine DCE and TTO data and had good model performance. This method is appealing statistically since it makes better use of the data. Method (4) used individual level data whereas method (3) used mean level data, and so method (4) does not suffer from problems associated with having only 10 or 19 data points.

Method (3) on mapping has the advantage that it has the lowest MAD from TTO values and is the easiest to implement. However it must be remembered that method (4) on the hybrid approach is estimated on individual level data that is typically noisier than mean level data, meaning it may not be expected to have as accurate predictions of mean TTO utilities as the mapping method that is estimated using mean level data. Confidence intervals for method (4) would directly reflect the numbers of observations in each model, whereas for method (3) bootstrapping could be used to calculate the confidence intervals to produce more reliable estimates. In addition the hybrid approach does not aim to predict the TTO data, but is by definition a hybrid of the DCE and TTO data, and hence is not necessarily expected to perform well using this criteria. Another finding was that the likelihood model performed better than the Bayesian method across all samples. This hybrid method is one of the methods being used in the new Euroqol Group's valuation protocol for the EQ-5D-5L, since it combines the theoretical advantages of TTO with the greater refinement of DCE, particularly at the upper end of the scale (29).

There is no agreed gold standard method for valuing health states in the field. In this study TTO has been used as a method for anchoring DCE values, but this evidence is only relevant where the aim is to predict TTO values.

An important difference between DCE and TTO tasks is that the DCE task makes no reference to duration, so it is assumed that the value of a state is independent of duration. The distributional assumptions underlying the DCE values are another source of divergence with TTO values. MacFadden (1974) (30) and others developed random utility theory that postulates that respondents choices between options A and B in a pair depend on their respective utility plus a random component. The distribution of the error term is used to estimate the latent utility function and individual deviation from the population mean utility values is treated as an analogue of individual error.

Limitations include the design of the DCE study. The design used a limited number of pairwise comparisons and was based upon the Huber and Zwerina design criteria (18) which, although widely used, do not guarantee optimal designs and on occasion cannot be used to estimate all the main effects of interest. This may have been one reason for some of the non-monotonicity. More sophisticated approaches to DCE study design using optimal and near optimal designs are now being recognised and applied in a health care context (31, 32) and some software can also allocate pairs into questionnaire versions rather than allocating pairs randomly to questionnaire versions (as undertaken here) where respondents may not see a mixture of levels for each dimension. However, a better DCE design is not likely to alter the results of the comparison of different anchoring methods, except that it may improve them to some degree. Method (1) anchored the results using a dead dummy that were based on two questions comparing either a severe health state or the worst health state to 'immediate death'. The limited number of states compared to 'immediate death' may have impacted on the results, but this is unlikely given that the descriptive system does not describe many states that may be thought of as being worse than 'immediate death', as indicated by the TTO value of the worst state.

Another potential limitation is the use of a condition-specific preference-based measure in this study (see (33) for an overview of condition-specific preference-based measures). One important question is whether the same relationships would hold for different classification systems. In addition, different results may have been found for the same classification system using a different DCE design and/or a different population, for example the non-monotonicity of the pollution dimension may be due to both the preferences of the general population in particular and/or the study design. The use of a condition-specific preference-based measure may mean that these results are not generalizable to other measures, although given that these methods rely on the close relationship between TTO and DCE it is

probable that similar findings would be found for other measures. In addition preliminary results from valuation work of the EQ-5D-5L suggests that the hybrid method works well (29). Research recently completed developing and valuing a generic social care outcome measure (ASCOT) with BWS has also found the mapping method to work well, though it has not been compared to other methods (28). In addition the upper anchor of 1 at full health may be questioned. All states were valued against an upper anchor of measure-specific full health, state 11111, and whether this is the same as generic full health and whether respondents understood this as equivalent to generic full health is debatable.

A potential limitation of the dataset used here is that respondents who valued states using both TTO and DCE (n=168) always saw the TTO elicitation method first. However, a previous study on this dataset found no differences between the DCE values that were elicited from respondents who had previously completed TTO tasks and from respondents who had not previously completed TTO tasks, suggesting that this should not have impacted on the results reported here (7). The response rate for the DCE postal survey for respondents who had not previously completed the DCE task is relatively low, at 24%, and it is possible that these respondents may differ from the non-responders, as it may be that responders had a special interest in the topic. As reported elsewhere the DCE and TTO samples are similar in the proportion of females with the DCE sample having a slightly older sample (7).

Ordinal methods such as discrete choice experiment (DCE) are an alternative for valuing health state classification systems as they are arguably easier to understand than commonly used cardinal methods of TTO and SG. However ordinal data has not been widely used to date due to the challenge of anchoring ordinal data onto the 1-0 full health-dead QALY scale. This paper explored two new methods for anchoring ordinal DCE data onto the 1-0 full health-dead QALY scale using mapping and estimation of a hybrid DCE and TTO model.

Both approaches required TTO data, but both predicted TTO data well using TTO data on just 10 health states. These new methods potentially have a useful role in producing values on the QALY scale using both ordinal and cardinal data that makes best use of the desirable properties of each elicitation technique and elicited data.

FUNDING

Data collection was funded by Novartis.

Table 1 Classification system of asthma-specific measure AQL-5D

Concern
1. Feel concerned about having asthma none of the time
2. Feel concerned about having asthma a little or hardly any of the time
3. Feel concerned about having asthma some of the time
4. Feel concerned about having asthma most of the time
5. Feel concerned about having asthma all of the time

Short of breath
1. Feel short of breath as a result of asthma none of the time
2. Feel short of breath as a result of asthma a little or hardly any of the time
3. Feel short of breath as a result of asthma some of the time
4. Feel short of breath as a result of asthma most of the time
5. Feel short of breath as a result of asthma all of the time

Weather and pollution
1. Experience asthma symptoms as a result of air pollution none of the time
2. Experience asthma symptoms as a result of air pollution a little or hardly any of the time
3. Experience asthma symptoms as a result of air pollution some of the time
4. Experience asthma symptoms as a result of air pollution most of the time
5. Experience asthma symptoms as a result of air pollution all of the time

Sleep
1. Asthma interferes with getting a good night's sleep none of the time
2. Asthma interferes with getting a good night's sleep a little or hardly any of the time
3. Asthma interferes with getting a good night's sleep some of the time
4. Asthma interferes with getting a good night's sleep most of the time
5. Asthma interferes with getting a good night's sleep all of the time

Activities
1. Overall, not at all limited with all the activities done
2. Overall, a little limitation with all the activities done
3. Overall, moderate or some limitation with all the activities done
4. Overall, extremely or very limited with all the activities done
5. Overall, totally limited with all the activities done

Table 2 Characteristics of respondents in valuation surveys

	TTO interview sample (n=307)	DCE postal survey (n=263)	P-value
Age:			
18-25	11.1%	3.4%	<0.001
26-35	18.6%	13.3%	
36-45	19.9%	17.1%	
46-55	16.3%	21.3%	
56-65	14.7%	24.3%	
>66	19.5%	20.5%	
Female	54.7%	56.3%	0.736
Have asthma	30.6%	30.3%	0.241
Self-reported EQ-5D scores:			
Male, female	0.84, 0.82	0.81, 0.82	0.302, 0.901

Note: Fisher's Exact test has been used to generate the p-values for female and having asthma, a t-test has been used to generate the p-values for EQ-5D scores by gender.

Table 3 Anchor based methods (1) to (2) – TTO and normalised DCE model estimates

	TTO	Latent DCE estimates	Method (1)	Method (2)
concern2	-0.028	0.053	0.012	0.008
concern3	-0.044*	-0.104	-0.024	-0.015
concern4	-0.054*	-0.394*	-0.099*	-0.058*
concern5	-0.081*	-0.649*	-0.139*	-0.096*
breath2	0.000	0.173	0.025	0.025
breath3	-0.036*	-0.017	-0.008	-0.003
breath4	-0.101*	-0.387*	-0.116*	-0.057*
breath5	-0.116*	-0.632*	-0.138*	-0.093*
pollution2	-0.019	0.375*	0.084*	0.055*
pollution3	-0.050*	0.067	-0.002	0.010
pollution4	-0.058*	-0.153	-0.051*	-0.023
pollution5	-0.121*	-0.427*	-0.085*	-0.063*
sleep2	0.018	-0.182	-0.022	-0.027
sleep3	0.010	-0.318*	-0.072*	-0.047*
sleep4	-0.033*	-0.636*	-0.125*	-0.094*
sleep5	-0.054*	-0.681*	-0.149*	-0.100*
activity2	-0.039*	-0.218*	-0.056*	-0.032*
activity3	-0.059*	-0.500*	-0.113*	-0.074*
activity4	-0.175*	-1.076*	-0.247*	-0.158*
activity5	-0.197*	-1.476*	-0.335*	-0.217*
Dead dummy			-1.000*	
Number of observations	2456	1559	2077	1559
Number of individuals	307	263	263	263
Inconsistencies	2	4	3	4
No. predictions >0.05 from observed TTO	19		34	24
No. predictions >0.1 from observed TTO	9		24	11
MAD from TTO	0.056		0.093	0.075
RMSD from TTO	0.070		0.118	0.093

Notes: *statistically significant at 5% level

Table 4 Method (3) - Mapping DCE onto TTO

	Method (3a) 10 states	Method (3b) 19 states	Method (3c) All states
DCE estimate	-0.142*	-0.141*	-0.118*
Constant	0.916*	0.928*	0.897*
Observations	10	19	99
R-squared	0.97	0.85	0.63
No. predictions >0.05 from observed TTO	50	52	43
No. predictions >0.1 from observed TTO	16	14	13
MAD from TTO	0.057	0.056	0.054
RMSD from TTO	0.072	0.071	0.069

Note: * statistically significant at 5% level.

Table 5 Method (4): DCE and TTO hybrid models

	Likelihood method			Bayesian method[1]		
	Method (4a) 10 states	Method (4b) 19 states	Method (4c) All states	Method (4a) 10 states	Method (4b) 19 states	Method (4c) All states
concern2	-0.008	-0.012	-0.023	0.008	0.003	-0.022
concern3	-0.029	-0.036	-0.029	-0.006	-0.010	-0.028
concern4	-0.069*	-0.070*	-0.064*	-0.052*	-0.056*	-0.069
concern5	-0.113*	-0.113*	-0.101*	-0.129*	-0.129*	-0.117
breath2	0.033	0.020	0.010	0.030	0.032	0.011*
breath3	-0.002	-0.008	-0.019	0.017	0.012	-0.016
breath4	-0.066*	-0.073*	-0.090*	-0.072*	-0.072*	-0.083
breath5	-0.092*	-0.100*	-0.094*	-0.081*	-0.089*	-0.100
pollution2	0.051	0.051	0.009	0.053*	0.058*	0.010*
pollution3	0.007	0.004	-0.023	-0.002	0.001	-0.027
pollution4	-0.034	-0.040*	-0.056*	-0.042*	-0.043*	-0.057
pollution5	-0.072*	-0.069*	-0.100*	-0.098*	-0.090*	-0.111
sleep2	-0.032	-0.033	-0.027	-0.022*	-0.017*	0.001*
sleep3	-0.053*	-0.052*	-0.048*	-0.068*	-0.059*	-0.046
sleep4	-0.102*	-0.096*	-0.085*	-0.118*	-0.110*	-0.081
sleep5	-0.106*	-0.100*	-0.105*	-0.124*	-0.114*	-0.097
activity2	-0.042*	-0.049*	-0.028	-0.027*	-0.039*	-0.035
activity3	-0.094*	-0.099*	-0.068*	-0.072*	-0.082*	-0.070
activity4	-0.171*	-0.180*	-0.167*	-0.209*	-0.209*	-0.185
activity5	-0.234*	-0.241*	-0.210*	-0.239*	-0.244*	-0.224
Teta	9.394	9.320	9.592	15.154	15.270	16.280
Number of observations	2055	2263	4015	2055	2263	4015
Inconsistencies	3	3	2	4	5	3
No. predictions >0.05 from observed TTO	48	50	43	56	51	46
No. predictions >0.1 from observed TTO	22	22	13	22	20	14
MAD from TTO	0.062	0.059	0.052	0.067	0.066	0.055
RMSD from TTO	0.080	0.067	0.067	0.084	0.083	0.070

Note: * statistically significant at 5% level.

[1] Significance (*) has been generated using 95% credible intervals, where if the credible interval does not include zero the coefficient is deemed significant at the 5% level.

Figure 1: DCE task

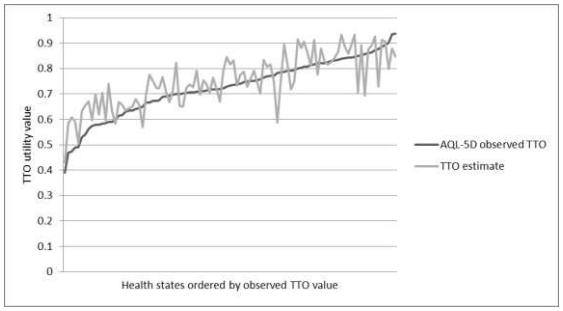
Health State A	Health State B
Feel concerned about having asthma **none of the time**	Feel concerned about having asthma **all of the time**
Feel short of breath as a result of asthma **none of the time**	Feel short of breath as a result of asthma **a little or hardly any of the time**
Experience asthma symptoms as a result of air pollution **none of the time**	Experience asthma symptoms as a result of air pollution **most of the time**
Asthma interferes with getting a good night's sleep **all of the time**	Asthma interferes with getting a good night's sleep **a little or hardly any of the time**
Overall, **a little limitation** in every activity done	Overall, **moderate or some limitation** in any activity done

Which health state do you think is better? *(please tick one box only)*

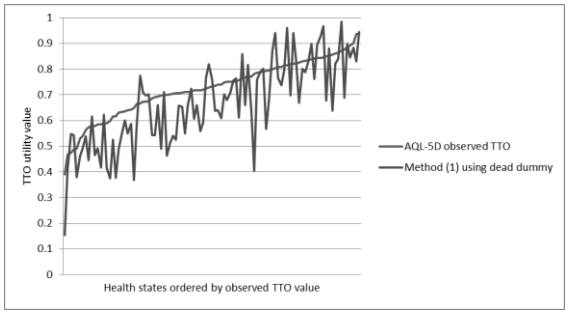
☐ ☐

Figure 2 Predicted utility and observed TTO

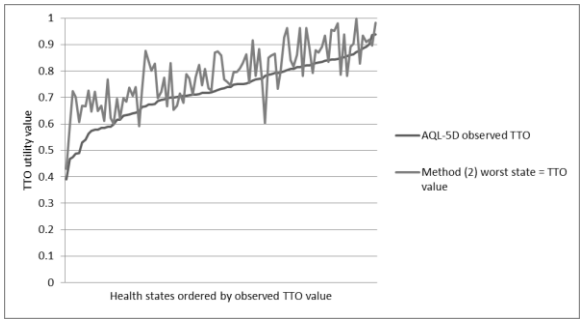
TTO estimates



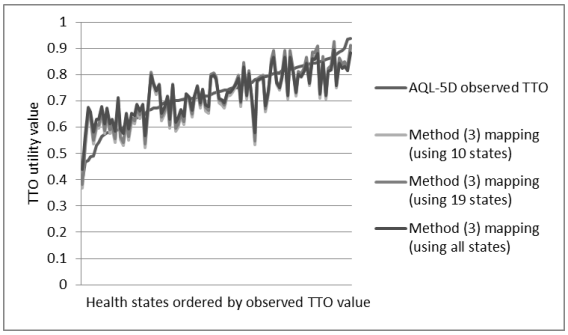
Method 1: anchored using dead dummy



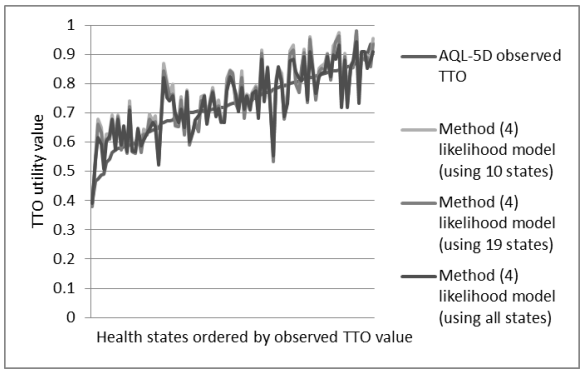
Method 2: anchored using worst state = TTO value



Method 3: mapping



Method 4: DCE-TTO hybrid estimates using likelihood model



Appendix 1

TIME BOARD 1

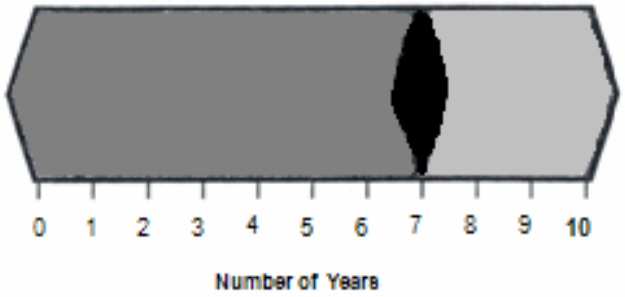
Feel concerned about having asthma none of the time

Feel short of breath as a result of asthma none of the time

Experience asthma symptoms as a result of air pollution none of the time

Asthma interferes with getting a good night's sleep none of the time

Overall, not at all limited with all the activities done



LIFE A

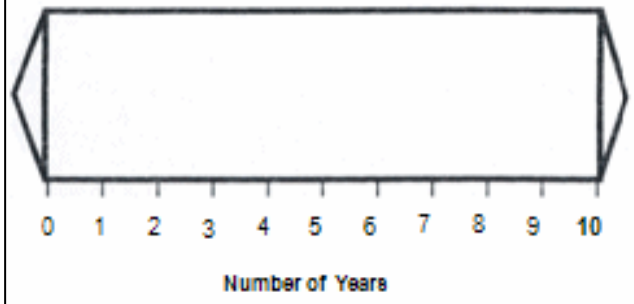
Feel concerned about having asthma all of the time

Feel short of breath as a result of asthma all of the time

Experience asthma symptoms as a result of air pollution all of the time

Asthma interferes with getting a good night's sleep all of the time

Overall, totally limited with all the activities done



LIFE B

Appendix 2

1. A combined likelihood function

We may combine the data from the TTO and DCE datasets as follows where *v* represents utility and x_{ij} is a vector of dummy variables for each level λ of dimension ∂ of the health state classification system where level $\lambda = 1$ is the baseline for each dimension. Health state j=1,2 …N is valued by respondent i=1,2…N^i. For the linear regression part we assume a normal distributed error leading to:

$$v_i = \alpha + \sum_{j=1}^{nd} \beta_j x_{ij} + e_i = \alpha + \beta' x + e_i$$

$$e_i \sim N(0, \sigma^2)$$

This can be rewritten as:

$$f(v_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{\left(v_i - \sum_{j=1}^{nd} \beta_j x_{ij} \right)^2}{2\sigma^2} \right)$$

and leading to the log likelihood function:

$$\text{loglik} = \log\left(\prod_{i=1}^{N} f(v_i) \right) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \sum_{i=1}^{N} \frac{\left(v_i - \sum_{j=1}^{nd} \beta_j x_{ij} \right)^2}{2\sigma^2}$$

For the discrete choice data we may say:

$$P(\text{left} > \text{right}) = P(v_l) > P(v_r)$$

$$v_l = \sum_{j=1}^{nd} \beta_j x_{lj} + e_l; \quad v_r = \sum_{j=1}^{nd} \beta_j x_{rj} + e_r$$

$$P(\text{left} > \text{right}) == \frac{1}{\left(1 + \exp\left(-\sum_{j=1}^{nd} \beta_j \left(x_{lj} - x_{rj}\right)\right)\right)}$$

$$P(\text{right} > \text{left}) = \frac{\exp\left(-\sum_{j=1}^{nd} \beta_j \left(x_{lj} - x_{rj}\right)\right)}{\left(1 + \exp\left(-\sum_{j=1}^{nd} \beta_j \left(x_{lj} - x_{rj}\right)\right)\right)}$$

$$\text{Loglik} = \sum_{i=1}^{Npair} N^i_{LGTR} \log\left(\frac{1}{\left(1 + \exp\left(-\beta' \Delta x_i\right)\right)}\right) + \sum_{i=1}^{Npair} N^i_{RGTL} \log\left(\frac{\exp\left(-\beta' \Delta x_i\right)}{\left(1 + \exp\left(-\beta' \Delta x_i\right)\right)}\right)$$

Where Δx^j is a vector measuring the difference in the dummy variables that characterise the health states in comparison *j*. The combination of the two may be seen as a simple product while acknowledging that they may differ up to a constant. The following likelihood was used to combine both sets of data:

$$\text{loglik} = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \sum_{i=1}^{N} \frac{\left(v_i - \sum_{j=1}^{nd} \beta_j x_{ij}\right)^2}{2\sigma^2} +$$

$$\sum_{i=1}^{Npair} N^i_{LGTR} \log\left(\frac{1}{\left(1 + \exp\left(-\theta\beta' \Delta x_i\right)\right)}\right) + \sum_{i=1}^{Npair} N^i_{RGTL} \log\left(\frac{\exp\left(-\theta\beta' \Delta x_i\right)}{\left(1 + \exp\left(-\theta\beta' \Delta x_i\right)\right)}\right)$$

2. A Bayesian approach

Methods 1-3 in the paper use random effects models and force the constant term to 1 (or zero). To compare these results to the results of hybrid method (4) we have to redefine the likelihoods, and here it is done using a Bayesian approach.

In the logistic model, used for the DCE data, we assume:

$$c_i^j \sim \text{Bernouilli}\left(p_i^j\right) \quad i = 1,..,N_{subjects}, \quad j = 1,...,N_{states}^i$$
$$\text{logit}\,(p_i^j) = \beta_i' \Delta x^j \quad i = 1,..,N_{subjects}, \quad j = 1,...,N_{states}^i$$
$$\beta_i \sim N(\bar{\beta}, \delta) \quad i = 1,..,N_{subjects}$$

Here, c_j^i is the answer of individual *i* to a discrete choice *j* (between two states), and β_i is a subject specific vector of parameters weighing the differences between the health states. Finally, $\bar{\beta}$ is the vector of average weights which is the main focus here.

In the linear model used for the TTO data, where v_i^j is the TTO value given by individual *i* to state *j*, we assume:

$$v_i^j \sim N(\beta_i' \Delta x^j, \sigma^i) \quad i = 1,..,N_{subjects}, \quad j = 1,...,N_{states}^i$$
$$\beta_i \sim N(\bar{\beta}, \delta) \quad i = 1,..,N_{subjects}$$
$$\sigma^i \sim \Gamma(g_1, g_2) \quad i = 1,..,N_{subjects}$$

In the hybrid approach we are using the same formulae as in the state approaches. However, we are saying that the mean beta's in both approaches are similar except for a constant ϑ. So, the whole model is now:

$$c_i^j \sim \text{Bernouilli}\left(p_i^j\right) \quad i = 1,..,N_{subjects}, \quad j = 1,...,N_{states}^i$$
$$\text{logit}\,(p_i^j) = \beta_i^{DCE\,\prime} \Delta x^j \quad i = 1,..,N_{subjects}^{DCE}, \quad j = 1,...,N_{states}^i$$
$$\beta_i^{DCE} \sim N(\bar{\beta}^{DCE}, \delta) \quad i = 1,..,N_{subjects}^{DCE}$$
$$v_i^j \sim N(\beta_i^{TTO\,\prime} \Delta x^j, \sigma^i) \quad i = 1,..,N_{subjects}^{TTO}, \quad j = 1,...,N_{states}^i$$
$$\beta_i^{TTO} \sim N(\vartheta \bar{\beta}^{DCE}, \delta) \quad i = 1,..,N_{subjects}^{TTO}$$
$$\sigma^i \sim \Gamma(g_1, g_2) \quad i = 1,..,N_{subjects}$$

Reference List

(1) Pliskin JS, Shepard DS, Weinstein MC. Utility functions for life years and health states. Operations Research 1980;28(1):206-54.

(2) Bleichrodt H, Johannesson M. Standard gamble, time trade-off and rating scale: Experimental results on the ranking properties of QALYs. Journal of Health Economics 1997;16(2):155-75.

(3) Coast J, Flynn T, Natarajan L, et al. Valuing the ICECAP capability index for older people. Social Science and Medicine 2008;67(5):874-82.

(4) Netten A, Burge P, Malley J, et al. Outcomes of social care for adults: developing a preference weighted measure. Health Technology Assessment 2012;16(16):1-153.

(5) Burr JM, Kilonzo M, Vale L, Ryan M. Developing a Preference-Based Glaucoma Utility Index Using a Discrete Choice Experiment. Optometry and Vision Science 2007;84(8):E797-E809.

(6) Hakim Z, Pathak DS. Modelling the EuroQol data: a comparison of discrete choice conjoint and conditional preference modelling. Health Economics 1999;8:103-16.

(7) Brazier J, Rowen D, Yang Y, Tsuchiya A. Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale. Eur J Health Econ 2012;13(5):575-87.

(8) Osman LM, McKenzie L, Cairns J, et al. Patient weighting of importance of asthma symptoms. Thorax 2001;56:138-42.

(9) Ryan M, Netten A, Skatun D, Smith P. Using discrete choice experiments to estimate a preference-based measure of outcome--an application to social care for older people. Journal of Health Economics 2006;25(5):927-44.

(10) Ratcliffe J, Brazier J, Tsuchiya A, Symonds T, Brown M. Using DCE and ranking data to estimate cardinal values for health states for deriving a preference-based single index from the sexual quality of life questionnaire. Health Economics 2009;18(11):1261-76.

(11) Flynn T, Louviere JJ, Marley AAJ, Coast J, Peters TJ. Rescaling quality of life values from discrete choice experiemtns for use as QALYs: a cautionary tale. Population Health Metrics 2008;6:6.

(12) Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate societal health state utility values. Journal of Health Economics 2012;31(1):306-18.

(13) Norman R, Viney RC, Brazier J, et al. Valuing SF-6D health states using a Discrete Choice Experiment. Medical Decision Making 2013;In Press.

(14) Young T, Yang Y, Brazier J, Tsuchiya A. The Use of Rasch Analysis in Reducing a Large Condition-Specific Instrument for Preference Valuation: The Case of Moving from AQLQ to AQL-5D. Medical Decision Making 2011;31(1):195-210.

(15) Juniper EF, Guyatt GH, Ferrie P, Griffith L. Measuring quality of life in asthma. American Journal of Respiratory Disease 1993;147:832-8.

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

(17) Gudex C. Time Trade-Off User Manual: Props and Self-Completion Methods. University of York: Centre for Health Economics; 1994.

(18) Huber J, Zwerina K. The importance of utility balance in efficient choice designs. Journal of Marketing Research 1996;33:307-17.

(19) Lancsar E, Louviere JJ. Conducting Discrete Choice Experiments to Inform Healthcare Decision Making. PharmacoEconomics 26, 661-677. 2008.
Ref Type: Journal (Full)

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

(21) Yang Y, Brazier J, Tsuchiya A, Young T. Estimating a Preference-Based Index for a 5-Dimensional Health State Classification for Asthma Derived From the Asthma Quality of Life Questionnaire. Medical Decision Making 2011;31(2):281-91.

(22) McCabe C, Brazier J, Gilks P, et al. Using rank data to estimate health state utility models. Journal of Health Economics 2006;25(3):418-31.

(23) Salomon JA. Reconsidering the use of rankings in the valuation of health states: a model for estimating cardinal values from ordinal data. Popul Health Metr 2003;1(1):12.

(24) 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. Eur J Health Econ 2010;11(2):215-25.

(25) Birnbaum MH, Parducci A, Gifford RK. Contextual effects in information integration. Journal of Experimental Psychology 88[2], 158-170. 1971.
Ref Type: Journal (Full)

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

(27) Ratcliffe J, Couzner L, Flynn T, et al. Valuing Child Health Utility 9D Health States with a Young Adolescent Sample: A Feasibility Study to Compare Best-Worst Scaling Discrete-Choice Experiment, Standard Gamble and Time Trade-Off Methods. Applied Health Economics and Health Policy 2011;9(1).

(28) Netten A, Burge P, Malley J, et al. Outcomes of social care for adults: developing a preference-weighted measure. Health Technology Assessment 2012;16:1-165.

(29) Devlin N, Shah K, Mulhern B, Feng Y, Tsuchiya A, van Hout BA. Developing an EQ-5D-5L value set for England. Health Economics Study Group, Sheffield, January 2014 . 2014.
Ref Type: Journal (Full)

(30) McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, ed. Frontiers in Econometrics.New York: Academic Press; 1974. p. 105-42.

(31) Viney R, Savage E, Louviere JJ. Empirical investigation of experimental design properties of discrete choice experiments in health care. Health Economics 2005;14:349-62.

(32) Street D, Burgess L. The Construction of Optimal Stated Choice Experiments: Theory and Methods. Hoboken, New Jersey: Wiley; 2007.

(33) Brazier JE, Rowen D, Mavranezouli I, et al. Developing and testing methods for deriving preference-based measures of health from condition specific measures (and other patient based measures of outcome). Health Technology Assessment 2012;16(32).