Work in progress – not for citation or quotation

Paper prepared for the 69th meeting of the Health Economists Study Group, University of York, 26-29 June 2006

USING RANK AND DISCRETE CHOICE DATA TO ESTIMATE HEALTH STATE UTILITY VALUES: THE CASE OF THE AQL-5D

John Brazier, Yaling Yang and Aki Tsuchiya
Health Economics and Decision Science, School of Health and Related Research, The University of Sheffield

Abstract

There is increasing interest in using ordinal methods for eliciting health state utility values as an alternative to conventional, cardinal, methods such as standard gamble and time trade-off. This study shows how two different kinds of ordinal data, rank and DCE, can be used to generate health state utility values for calculating QALYs. It presents results from modelling TTO, ranking, and DCE data to value an asthma specific health state descriptive system (the AQL-5D). The rank data generated similar values to TTO, but DCE generated a wider range of values. Rank and DCE data offer promising alternatives to conventional cardinal methods of SG and TTO, but there is a large and important research agenda to address.

Acknowledgements

This study is funded by Novartis. John Brazier is funded by the Medical Research Council Health Services Research Collaboration. We are grateful to Jason Leman and Tracey for their input and to all the participants who took part. The usual disclaimer applies.

Corresponding author: John Brazier, HEDS, ScHARR, University of Sheffield, Regent Court, 30 Regent Street. j.e.brazier@shef.ac.uk.

1.0 Introduction

The status of preference-based measures was considerably enhanced by the recommendations of the U.S. Public Health Service Panel on Cost-Effectiveness in Health and Medicine to use them in economic evaluation (Gold et al, 1996). The use of preference-based measures of health for calculating Quality Adjusted Life Years (QALYs) has grown considerably over the last decade with the increasing use of economic evaluation to inform health policy, for example through the establishment of bodies such as the National Institute of Health and Clinical Excellence in England and Wales (NICE, 2004).

To be a preference-based measure, the health state valuation technique must be choice-based (Drummond et al, 2005; NICE, 2004; Gold et al, 1996). The two main choice-based techniques used to value preference-based measures have been the cardinal methods of standard gamble (SG) and time trade-off (TTO) (e.g. Dolan, 1997; Brazier et al, 2002; Feeny et al, 2002). However, there are concerns about these cardinal methods given that they are likely to be contaminated by factors other than a respondent's preference for the state, such as risk aversion in the case of standard gamble, or time preference and aversion to losses for TTO (Bleichrodt, 2003). Furthermore, these tasks are cognitively complex and respondents might have some difficulty with them, particularly those in vulnerable groups such as the very elderly or children. For these reasons there has been increasing interest in using ordinal tasks that require the respondent to rank two or more states (Kind, 1982; Salomon, 2003; McCabe et al, 2006).

The ability to derive cardinal health state values from ordinal information comes from the assumption that a respondent's ordinal preferences over a set of states will be related to a latent variable. It allows for the fact that individuals make errors of judgement and sometimes may rank the health state with a lower value above another state with a higher value. The proportion of occasions on which such an error is made is related to the distance between the mean values of the states. There will be fewer errors and more agreement in ordinal preferences the further apart the mean values for two states in terms of the latent variable. This has been the basis for the more general use of discrete choice experiments. By making additional assumptions it is possible to ‘explode’ ranking data into discrete choice data, whereby the ordering of X states is essentially seen as a sequence of discrete choices.

This paper presents results from a study concerned with valuing a condition specific health state descriptive system derived from the Asthma Quality of Life Questionnaire (Juniper et al, 1993). It compares the results from using ranking and DCE data with TTO in a sample of the UK general population. It presents a method for generating cardinal health state values from rank and DCE data that lie on the full health-death scale required to generate QALYs. It then compares the results obtained from TTO, ranking and DCE. Finally, it examines whether the results from the DCE data obtained from a sample that had previously been interviewed using ranking and TTO are similar to those obtained from a ‘cold’ sample that had not previously been interviewed. In the spirit of HESG, this is very much work in progress and in the discussion we consider the implications of the work and a research agenda in this area.

2.0 The theoretical basis for deriving cardinal values for health states from ordinal information

The idea of obtaining cardinal values from ordinal data first came from the work of Thurstone (1927) who proposed the ‘law of comparative judgement’. This was first recognised by Fanshel and Bush (1970) as a offering a method for deriving cardinal utility values for health states from rank data and later taken up by Kind (1982) who scored the sleep dimension of the NHP and more recently the EQ-5D classification (Kind, 2005).

The law of comparative judgement considers the proportion of times that health state *A* is considered worse than health state *B*. The preferences over the health states represent a latent cardinal utility function. Individual’s stated preferences draw upon this latent function but in an imperfect manner; i.e. there are errors in individual’s expression of the latent utility function. The closer two health states A and B lie on the latent utility function the greater the likelihood that an individual will incorrectly state that they prefer B to A, when in fact the utility they expect to gain from health state A is greater than the utility they expect to gain from health state B. Thus, at the aggregate level, there will be a relationship between observed ordinal preferences and the underlying cardinal latent utility function.

Thurstone’s approach has been modified in a number of ways, including the one that uses a logistic function (Luce, 1959; McFadden, 1974). It uses a conditional logistic regression model as a means of modelling this latent utility function from ordinal data. The details of this model are specified below. Another important modification in this context is that in modelling a population level latent utility function from individual rank data, the error is being characterised in terms of the deviation of the individuals’ preferences from the population preferences; i.e. variation in individual preferences within a population is considered analogous to Thurstone’s individual level perceptual error.

The models described above are concerned with pair wise data. To use rank data requires an additional assumption in order to explode the rank data into a series of pair wise choices or more commonly that the states are selected one at a time from a set, with the preferred state being chosen first, followed by the second state and so on. This is made possible by the assumption of independence from irrelevant alternatives (IIA): the ordering of a pair of states does not depend on the other states being considered.

Recently Salomon (2003) presented work that applied conditional logistic regression models to the rank data collected as part of the York Measurement and Valuation of Health Study (MVH) and McCabe, Brazier, Tsuchiya and others (2006) to the SF-6D and HUI2. The rank model does not produce utilities on the 0-1 scale necessary for use in estimating QALYs, because the rank data on health states alone does not enable the anchoring of the values to 0 for dead. For this reason, the values generated by the logit model were transformed onto the full health-death scale needed to generate QALYs by Salomon (2003) and McCabe et al (2006) by including dead in the ranking data and rescaling or normalising the regression coefficients so that being dead achieves a predicted value of zero.

Experience to date of using conditional logit to model ordinal data to generate health state values for calculating QALY has been limited to the work of Salomon (2003) and McCabe et al (2006) who looked at rank data. DCE has been a widely used tool in health economics, but its use has been limited mainly to valuing non-health outcomes and process benefits rather than trying to value health benefits for deriving QALYs . A limited number of studies have used DCE to value health states for there own sake (Hakim and Pathak, 1999; Johnson et al, 2000; McKenzie et al, 2001; Ryan et al, 2006), but none have made the link to the full-health death scale required for generating QALYs. Ratcliffe and Brazier (2006) have attempted a partial solution by normalising the DCE results using the TTO value for the 'pits' state, or the worst possible state. The study presented in this paper attempts to undertake a normalisation of DCE results around dead and to compare with TTO and rank results.

3.0 Methods

3.1 Asthma Quality of Life Questionnaire

The paper presents results from valuing a health state descriptive system derived from the Asthma Quality of Life Questionnaire (AQLQ). The AQLQ consists of 32 items with 7 levels each, covering 4 dimensions: symptoms (12 items), activity limitations (11 items), emotional function (5 items) and environmental stimuli (4 items). This is too large to be amenable to valuation. Therefore, based on the application of Rasch analysis and conventional psychometric tests, the AQLQ has been reduced to a 5-dimension health state classification system which we call AQL-5D (see an earlier HESG paper by Young et al, 2005). The dimensions of AQL-5D are: concern about asthma, shortness of breath, weather and pollution stimuli, sleep impact and activity limitations (Table 1). Each dimension has 5 levels of severity with level 1 denoting no problem and level 5 indicating extreme problem. By selecting one level for each dimension it is possible to define 3125 health states.

3.2 Interview

As described in a paper presented at the last HESG (Yang et al, 2006) the purpose of the study was to elicit values for a sample of AQL-5D states from a representative sample of the general public (target 300). Adults who consented to participate were interviewed in their own home by an experienced interviewer trained by the authors of this paper. The interview began with a warm-up exercise to familiarize respondents to the AQL-5D. Those respondents who regarded themselves as having asthma were asked to describe their current state of health using the AQL-5D classification system. Those respondents who did not regard themselves as having asthma were invited to imagine somebody with asthma, and were then asked to describe that person's condition using the AQL-5D classification system. The first valuation task was to rank 7 intermediate AQL-5D states, full health (AQL-5D health state 11111), worst state defined by the AQL-5D ('pits' state 55555) and immediate death. They were given cards depicting each of these states and asked to rank them together. The method of administering the ranking task has been used for the EQ-5D and SF-6D valuation studies and it conventionally seen as a warm up task to the main cardinal task.

The next task was to value the 7 intermediate states and 'pits', with an upper anchor of the best AQL-5D state. The survey used the TTO-prop method developed by the York Measurement and Valuation Health Group, which uses a 'time board' as a visual aid (Gudex, 1994). The respondents were then asked a series of socio-demographic questions. Finally, they were asked about their willingness to participate in a postal survey.

The selection of health states for the interviews was determined by the specification of the model to be estimated. In this study, 98 health states were selected out of the 3125 possible health states for the TTO valuation. The selection was on the basis of a balanced design, which ensured that any dimension-level (level λ of dimension δ) had an equal chance of being combined with all levels of the other dimensions. These 98 states were stratified into severity groups based on their total level score across the dimensions (simply the sum of the levels), and then randomly allocated into 14 blocks, so that each block has 7 health states. This procedure ensured that each respondent, who were allocated one of the 14 blocks, received a set of states balanced in terms of severity and that each state is valued the same number of times apart from the worst possible state, or the 'pits' state, which is valued by all respondents.

3.3 Follow-up postal survey

A DCE questionnaire was posted to those respondents who consented to the postal survey after approximately four weeks of the interview, with reminders sent to non-responders at roughly 2 weeks after the initial posting of the questionnaire. As in the interview, respondents were first asked to describe either themselves (if they have asthma) or somebody else with asthma using the AQL-5D classification system. Then respondents were asked to indicate which state they preferred in each pair of states.

The selection of states for the postal DCE were determined independently from the 98 states for the interview by an application of programme developed by Huber and Zwerina (200?, need to find out) in the statistical package SAS. This obtains an optimal statistical design for DCE based on level balance, orthogonality, minimal overlap and utility balance. This reduces the number of pair wise comparisons to a manageable number for this survey. From the output of the programme, 12 pair wise comparisons were selected and these were randomly allocated to two versions of the questionnaire with 6 choices. Two additional pair wise comparisons were included of AQL-5D states compared to being dead, and the same pairs were used in the two versions of the questionnaire.

In order to ascertain the effect of the preceding interview on the postal survey, two groups of respondents were recruited. The first group, the 'warm' group consisted of those who took part in the interview and agreed to participate in the postal follow-up survey. The second, the 'cold' group consisted of newly sampled respondents, with no previous contact regarding this valuation study.

4.0 Modelling health state values

4.1 TTO

The data from the TTO valuation exercise of the interviews were analysed using a one way error components random effects model was specified which takes account of variation both within and between respondents (Brazier et al, 2002). The model is defined as:

$$Y_{ij} = f(\beta'\mathbf{x}_{ij}) + \epsilon_{ij} \qquad (2)$$

Where i=1,2 …n represent individual health state values and j = 1,2…m represents respondents. The dependant variable Y_{ij} is the disvalue (1-mean TTO value) for health state i valued by respondent j and $\boldsymbol{x}$ is a vector of dummy explanatory variables ($X_{\partial\lambda}$) for each level λ of dimension ∂ of the simplified SQOL classification. Level λ = 1 acts as a baseline for each dimension. ϵ_{ij} is the error term which is subdivided as follows:

$$\epsilon_{ij} = u_j + e_{ij} \qquad (3)$$

Where u_j is respondent specific variation and e_{ij} is an error term for the *i*th health state valuation of the *j*th individual, and this is assumed to be random across observations. Details of other models run on the TTO model are given in Yang et al (2006).

4.2 Ranking

The rank ordered logit model was used to analyse the ranking data of the interviews. This model is based upon the assumption that the respondent makes a series of selections from smaller and smaller groups. Thus in ranking the 10 health states we assume that the respondent chooses the most preferred state from the full set, then chooses the most preferred state from the remaining 9 etc until all health states have been assigned a rank between 1 and 10. The independence of irrelevant alternatives assumption is required to characterise this process as equivalent to a series of pair wise choices i.e. the ranking of the pair is not affected by the other states that are ranked in the same exercise (Luce, 1959).

The rank ordered logit model states that respondent *i* has a latent utility function for state *j*, U_{ij} and given the choice of two states *j* and *k*, the respondent will choose state *j* over state *k* if $U_{ij} > U_{ik}$.

The expected value of each unobserved utility was assumed to be a linear function of the categorical levels on the dimensions of the AQL-5D. Following the approach taken by Salomon (2004) and McCabe et al (2006), the general model specification for each individual's cardinal utility function for state *j* is $U_{ij} = \mu_j + \epsilon_{ij}$ where μ_j is representative of the tastes of the population and ϵ_{ij} represents the particular tastes of the individual. If the error term ϵ has an extreme value distribution, then the odds of choosing state *j* over state *k* are $exp\{\mu_{j} - \mu_k\}$.

The general model specification for analysis of the ranking data is:

$$\mu_{ij} = f(\beta'\mathbf{x}_{ij} + \Phi D + u_{ij}) \qquad (4)$$

where μ= utility; i=1,2,…n represents respondents and j = 1,2,….m represents health states. The functional form is again assumed to be linear. The vector of dummies is as for equation (2), with the addition of a dummy variable (*D*) for the state dead. For all health states other than dead D = 0. It is necessary to transform the model by rescaling the coefficients relating to the levels of each dimension by dividing each level coefficient by the coefficient relating to death (Solomon, 2004; McCabe et al, 2006).

4.3 DCE

The data from the DCE postal surveys were analysed using a random effects probit model, which takes account of the repeated measurement aspect of the data (whereby multiple responses are obtained from the same individual). Again an additive specification is used like that in equation (2). The coefficients were rescaled in the same way as the rank data by dividing each level coefficient by the coefficient relating to death. These analyses are undertaken for the 'warm' sample that was previously interviewed and the 'cold' sample that were not.

4.4 Comparison of models

It is not possible to compare the performance of these models in the conventional sense in terms of their ability to predict actual TTO values, since the models are based on different types of data. There is no reason why rank or DCE models should produce the same results as the TTO model, although it could be thought that since both are based on ordinal data, Rank and DCE should produce similar results.

Models can be compared in terms of the sign and ordering of their coefficients. The sign of the coefficients on the levels of each dimension are expected to be negative since they are all worse than the baseline (i.e. level 1). Furthermore, the levels in each dimension have a logical ordering, whereby more severe levels should have larger decrements. The number of inconsistencies between significant coefficients is compared between the models. Another test is the normality of the errors. For interest, we examine the relationship between model predictions and TTO actual values including the mean absolute difference, the root mean square of the difference and the proportions of differences greater than 0.05 and 0.1. Finally the pattern of the predictions is compared.

5.0 Results

5.1 The Data set

A sample of 307 members of the public (response rate 40%) in South Yorkshire was interviewed. They were all included in the final dataset for analysis. The description of the sample is shown in Table 1. Among the respondents, more than half are female, between 36 to 65 years old, married or living with partner, and experienced serious illness in their family. In this sample, 53 (17.3%) have asthma, 22.5% respondents have a degree or equivalent, and 45.6% respondents receive full-time education after age 17.

There were 2455 TTO health state valuations generated and 3039 states ranked by the respondents at their interview. The average number of TTO valuations per intermediate health state was 22 (range from 19 to 22) and the 'pits' state (AQL-5D state 55555) was valued by every respondent (n=307). The mean health state values ranged from 0.39 to 0.94 and generally have fairly large standard deviations (around 0.2 to 0.4). The distribution of the values was negatively skewed. Full results of the TTO valuation are presented in Yang et al (2006).

Of the 308 individuals interviewed, 291 (94.5%) agreed to participate in the postal follow-up survey, and 168 of them (57.7%) returned the questionnaire after one reminder, generating 1336 observed pair wise comparisons for the warm survey. From the cold survey, 95 (23%) returned the questionnaire after one reminder, generating 741 pair wise comparisons.

5.2 Modelling

The untransformed results from modelling the rank and DCE data are presented on Table 2. The dimension level coefficients reflect progressively worse levels of the AQL-5D in all cases, except for one small inconsistency in the rank model (between breath levels 3 and 4) and a rather more important one in the overall DCE model (weather/pollution levels 2 is in the wrong direction).

The TTO and transformed rank and DCE models are presented in Table 3. The TTO model produced negative coefficients (for predicting TTO) as expected and they all had the expected ordering within a dimension of AQL-5D. As before, the rank model produced one inconsistency. The DCE models for the pooled data (i.e. warm plus cold) produced three positive coefficients and one inconsistency between significant coefficients (between weather/pollution 2 to level 1) . The warm DCE model produced five positive (i.e. wrong) coefficients and the cold model just one, with both producing one inconsistency. The weather/pollution dimension seemed to cause most difficulty for the DCE models, with a suggestion that the levels of this dimension do not conform to the suggested ordering.

The differences in the models in terms of coefficients estimated for each dimension level is best seen in Figure 1. [note: figures lost from the copy sent by Aki.] The coefficients of the rank and TTO models are quite similar and follow an orderly pattern against the levels of the AQL-5D. The DCE model for the pooled data set reveal number of more marked differences. The most noticeable differences lie at the lower end of concern, breath and weather/pollution and the upper ends of sleep and activity. Level 2 for the dimensions of concern, breath and weather/pollution are all positive and in the wrong direction (and quite markedly so for weather/pollution). Sleep and activity have coefficients with the right sign, but they are much larger for levels 4 and 5.

The similarity of the rank and TTO models can be seen in the plot of predicted values against actual TTO in Figure 2. It can be seen how closely the TTO and rank predictions follow the actual mean TTO values, with mean absolute errors of 0.051 and 0.065 and mean errors around zero. By contrast, the DCE model had a set of points that follow a steeper path, with the upper end having higher values than those for TTO and lower values at the lower end. This pattern in reflected in the coefficients in the simple regressions between actual TTO and predictions across the three models. The TTO and rank predictions have very similar gradients where the DCE is considerably steeper. This is also reflected in the differences between the mean values for the worst AQL-5D health state of 0.390 for actual TTO and predictions of 0.457 for TTO, 0.405 for rank and 0.172 for DCE.

There is interest in comparing DCE models estimated from those who were interviewed and those approached cold. The models presented in Table 3 show similar coefficients and mean absolute differences form TTO. The warm model has rather more significant coefficients, but then there were more observations with this data set.

6.0 Discussion

This study has shown how DCE and rank data can be used to generate health state values on the full health-death scale required to generate QALYs. As would be expected, the TTO model best predicted TTO actual values, but then there is no reason to expect rank and DCE data to produce the same values. Perhaps more surprising is the way the rank model was actually very similar to the TTO and the DCE model was different. It might have been expected that the rank and DCE model would be more alike. The modelling of rank data essentially treated it as DCE data, and aside from the IIA assumptions, it is otherwise the same. This may suggest that the rank and DCE tasks generate different data and that the IIA assumption used in rank data needs to be relaxed. It may also reflect the fact that the ranking task preceded the TTO in the same interview, whereas the DCE data were collected after a period of 4 weeks in a postal survey.

Evidence from this survey suggested little difference between the cold and warm survey. It might be thought that respondents who were interviewed would give different values since they had more time to think about the condition or at least that their responses would have been influenced by the interview. However, the previous interview seemed to have little influence. This could have important implications for future applications of this approach.

The rank and DCE were really 'add-ons' to a study that was mainly designed to provide TTO valuations of the AQL-5D. As a consequence the rank data are collected in the same way as they were for the EQ-5D and SF-6D, where it is really a warm-up task rather than a task in its own right. It has been suggested in McCabe et al (2006) that perhaps the ranking exercise should be re-designed to make it more compatible with the conditional logit model that assumes a sequential process of selecting the states one at a time. Respondents could be given the states and asked to identify their preferred state, then out of the remainder the next best and so forth. The DCE data were collected using a postal method and this may have reduced the quality of data and it certainly resulted in a substantially lower response rate. Perhaps more importantly, the design of the DCE was far from ideal, with the pairs of AQL-5D states being selected by Huber and Zwerina (1996) programme, whereas there are better ways of doing this now. Finally, the comparisons with being dead should have been built into the original design when the states were selected for inclusion in the survey.

There are technical details about the need to use a conditional logit model for analysing DCE in a manner comparable with ranking data, but this should not substantively affect the results. There are more fundamental concerns with these types of models since they make restrictive distributional assumptions about the coefficients. Of particular concern is that some orderings are logically determined. For example, suppose there is a health state pair: i and j, and $\mu_i - \mu_j = X$, say 0.2, on the latent variable scale standardised to 1 for full health and 0 for dead. The current approach to modelling ordinal data assumes that any two states that are apart from each other by X will have the same proportion of respondents incorrectly ranking j over i. However, it is reasonable to assume that the probability of error will not only be a function of how apart the two states are, but also whether or not the two states have a logically determined ordering. Suppose there are two sets of health state pairs that are apart by X, where pair 1 has no logically determined ordering (e.g. 11122 and 33111) whereas pair 2 has a logically determined ordering (e.g. 11122 and 11133). It is reasonable to expect that the proportion of responses that rank j over i will be different across pair 1 and pair 2. This becomes particularly problematic when one of the states is full health or pits. This means that the structure of the error term in equation (4) needs to be more sophisticated than it currently is. There are now more advanced econometric modelling techniques known as mixed logit models that need to be explored with both these data sets (Train, 2003). This would also overcome the IIA assumption underlying the way rank data are being analysed.

DCE and ranking method depends on respondents making errors, but if everyone gets it right then there is a risk that the distance may be over stated. There was no evidence of this problem here. Another problem may arise where all respondents regard all the states as better than being dead. This does not seem to have been a problem for the AQL-5D, but may arise with systems describing more mild conditions.

This study has built on earlier work to show that rank and DCE data sets can be use to value health states for generating QALYs. However, further work it needed in order to develop this approach. At a special session at the last iHEA conference we raised the following research questions (Brazier and Tsuchiya, 2005):

1. Are DCE/Ranking methods really easier than TTO or SG?

A key attraction of the ranking and DCE task is the assumption that the task is easier. However, this is not the right question, since what is more relevant is whether it is easier for the collection of the data needed to value a descriptive system and it may be that a better and more valid DCE design would take longer to do. The states to be valued, the necessary sample sizes and methods of administration would differ between ordinal and cardinal methods. A true comparison of ease of use needs to be explored in a comparative study where each method is used within the same budget.

2. Does DCE/Ranking produce different estimates from TTO and SG?

There is no reason why DCE and rank methods should produce the same values as TTO or SG, indeed there are good reasons for supposing they will generate different values due to risk aversion, time preference and so on. McCabe et al (2006) found rank models produced similar results to SG for the SF-6D and HUI2, but Solomon (2003) and Ratciffe and Brazier (2006) found ranking produced different results from TTO after normalising being dead to zero. Our results suggest ranking produces similar results but DCE are different. Clearly more research is needed into the relationship between these different methods.

3. Does DCE better reflect preferences for health states than TTO or SG?

Random utility theory is well developed and offers an alternative to TTO and SG. It would seem to avoid the problems associated with these techniques, but more work is needed to understand the theoretical basis of the ordinal techniques for this application. In addition, whereas the literature on TTO and SG often consider excluding respondents because they violate logical consistency, DCE and ranking model individual behaviour taking into account that people make mistakes. Keeping in mind the point raised above regarding the more sophisticated error structures, these ordinal approaches seem to be a more realistic way of reflecting public preferences in decision making. However, there needs to be more empirical work into the validity of these methods.

4. Design and analytical issues

These include: which is the best format for these tasks (pair wise, three way or even best worst as suggested by Flynn et al)? How should the states and pairs be identified? Finally how should these data be analysed. There are more advanced econometric methods available that make fewer assumptions to be explored.

7.0 Conclusion

This study has shown how rank and DCE data can be used to generate health state values for calculating QALYs. The rank data generated similar values to TTO, but DCE may generate a wider range of values. Rank and DCE data offer promising alternatives to conventional cardinal methods of SG and TTO, but there is a large and important research agenda to address.

References:

Bleichrodt H (2002). A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Economics* 11(5):447-456.

Brazier JE, Deverill M, Harper R, Booth A (1999). A review of the use of health status measures in economic evaluation. *Health Technology and Assessment* 3 (9).

Brazier J, Ratcliffe J, Salomon J, Tsuchiya A (2006 forthcoming). *The measurement and valuation of health benefits for economic evaluation*. Oxford University Press, Oxford.

Brazier J, Roberts J, Deverill M. The estimation of a preference based single index measure for health from the SF-36 (2002). *Journal of Health Economics* 21: 271-292.

Brazier J, Tsuchiya A (2005), *Using Discrete Choice Experiments to estimate quality weights for Quality Adjusted Life Years: the issues*, presentated at the International Health Economics Association meeting, Barcelona summer 2005

Dolan P. Modelling valuation for Euroqol health states. *Medical Care* 1997;35:351-363.

Fanshel S, Bush JW (1970). A health status index and its application to health services outcomes. *Operations Research* 18(6):1021-1066.

Gudex C. *Time Trade-Off User Manual: Props and Self-Completion Methods*. Centre for Health Economics, 1994

Hakim Z, Pathak DS (1999). Modelling the EuroQol data: a comparison of discrete choice conjoint and conditional preference modelling. *Health Economics* 8(2): 103-116.

Juniper EF, Guyatt GH, Ferrie PJ, Griffith LE. Measuring quality of life in asthma. *American Review of Respiratory Disease* 1993;147:832-838.

Kind P (1982). A comparison of two models for scaling health indicators. *International Journal of Epidemiology* 11 (3):271–275.

Kind P (2005). Applying paired comparisons models to EQ-5D valuations – deriving TTO utilities from ordinal preferences data. In: *EQ-5D concepts and methods: a developmental history*. Kind P, Brooks R, Rabin R (eds): Springer, Netherlands. pp. 201-220.

Luce RD (1959) *Individual choice behavior: a theoretical analysis.* John Wiley & Sons, Inc., New York.

McCabe C, Brazier J, Gilks P, Tsuchiya A, Roberts J, O'Hagan A, Stevens K (2006). Using rank data to estimate health state utility models. *Journal of Health Economics* 25(3): 418-431.

McFadden D (1974). Conditional logit analysis of qualitative choice behavior. In *Frontiers in econometrics*, Zarembka P (ed): Academic Press, New York. pp. 105-142.

MVH Group (1995). *The measurement and valuation of health: Final report on the modelling of valuation tariffs*. Centre for Health Economics, University of York.

Ryan M, Farrar S (1994). *A pilot study of using conjoint analysis to establish the views of users in provision of orthodontic services in Grampian.* Health Economics Research Unit Discussion Paper 07/94. University of Aberdeen, Aberdeen.

Ratcliffe J, Brazier JE (2006). *Estimation of a preference based single index from the sexual quality of life questionnaire (SQOL) using ordinal data*. Health Economics and Decision Science Discussion Paper 06/06, ScHARR, University of Sheffield. http://www.sheffield.ac.uk/scharr/sections/heds/discussion.html.

Ryan M, Netten A, Skatun D, Smith P (2006). Using discrete choice experiments to estimate a preference-based measure of outcome – an application to social care for older people. *J Health Econ* (in press).

Salomon JA (2003). Reconsidering the use of rankings in the valuation of health states: a model for estimating cardinal values from ordinal data. *Popul Health Metr.* 1(1): 12. http://www.pophealthmetrics.com/content/1/1/12.

Salomon JA, Murray CJ (2004). A multi-method approach to measuring health state valuations. *Health Economics* 13(3): 281-290.

Train K (2003) Discrete choice methods with simulation. Cambridge University Press. Cambridge.

Yang Y, Tsuchiya A, Brazier J, Young T (2006), *Deriving a preference-based measure for health from the AQLQ*. Paper presented at the Health Economists' Study Group meeting, City University, January 2006

Young T, Yang Y, Brazier J, Tsuchiya A. *Using Rasch analysis to aid the construction of preference based measures from existing quality of life instruments.* Paper presented at the Health Economists' Study Group meeting, Newcastle, summer 2005

Table 1: Asthma quality of life classification (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 & 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: Models estimated from Rank data and DCE 'warm' and 'cold' sample

Dimension level	Rank coefficient (Logit)	DCE coefficient (probit)		
		All data	Warm data	Cold data
concern2	.264*	-0.047	-0.079	0.060
concern3	.585*	0.095	0.022	0.216
concern4	1.127*	0.385*	0.373*	0.469*
concern5	1.38*	0.542*	0.454*	0.743*
breath2	.434*	-0.096	-0.161	0.061
breath3	.675*	0.032	-0.015	0.135
breath4	.650*	0.451*	0.338*	0.710*
breath5	1.081*	0.535*	0.470*	0.695*
pollution2	.101	-0.328*	-0.395*	-0.233
pollution3	.387*	0.006	-0.014	0.017
pollution4	.464*	0.200*	0.180	0.249
pollution5	.619*	0.332*	0.349*	0.263
sleep2	.025	0.087	0.093	0.079
sleep3	.230	0.281	0.278*	0.363*
sleep4	.515*	0.484*	0.383*	0.742*
sleep5	.772*	0.580*	0.431*	0.897*
activity2	.470*	0.216*	0.233*	0.235
activity3	.757*	0.441*	0.423*	0.524*
activity4	1.409*	0.958*	0.963*	1.043*
activity5	1.800*	1.300*	1.343*	1.341*
N		2077	1336	741
Death dummy	9.496*	3.886*	3.681*	4.484*
Inconsistencies**	1	1	1	1

*statistically significant at 5% level

** relating to statistically significant dimension only

Table 3: Comparison of models estimated from TTO and rescaled DCE and Rank data

Dimension level	TTO	Rank (logit)	Discrete choice experiment		
			Full data	warm data	Cold data
concern2	-0.047*	-0.028*	0.012	0.021	-0.013
concern3	-0.064*	-0.062*	-0.024	-0.006	-0.048
concern4	-0.074*	-0.119*	-0.099*	-0.101*	-0.105*
concern5	-0.095*	-0.145*	-0.139*	-0.123*	-0.166*
breath2	-0.024	-0.046*	0.025	0.044	-0.014
breath3	-0.045*	-0.071*	-0.008	0.004	-0.030
breath4	-0.107*	-0.068*	-0.116*	-0.092*	-0.158*
breath5	-0.116*	-0.114*	-0.138*	-0.128*	-0.155*
pollution2	-0.017	-0.011	0.084*	0.107*	0.052
pollution3	-0.028	-0.041*	-0.002	0.004	-0.004
pollution4	-0.063*	-0.049*	-0.051*	-0.049	-0.056
pollution5	-0.099*	-0.065*	-0.085*	-0.095*	-0.059
sleep2	-0.013	-0.003	-0.022	-0.025	-0.018
sleep3	-0.029	-0.024*	-0.072*	-0.076*	-0.081*
sleep4	-0.054*	-0.054*	-0.125*	-0.104*	-0.165*
sleep5	-0.069*	-0.081*	-0.149*	-0.117*	-0.200*
activity2	-0.029	-0.049*	-0.056*	-0.063*	-0.052
activity3	-0.044*	-0.080*	-0.113*	-0.115*	-0.117*
activity4	-0.139*	-0.148*	-0.247*	-0.262*	-0.233*
activity5	-0.164*	-0.190*	-0.335*	-0.365*	-0.299*
Death dummy		-1.000*	-1.000*	-1.000*	-1.000*
N	2455	3039	2077	1336	741
Inconsistencies**	0	0	2	3	1
No.>0.05 from TTO	22	30	34	31	40
No.>0.1 from TTO	11	11	23	19	31
MAD from TTO	0.051	0.065	0.092	0.09	.121
RMSD from TTO	0.065	0.080	0.119	0.114	0.151
Mean Error	0.015	0.010	0.053	0.030	0.1

*statistically significant at 5% level
** relating to statistically significant dimensions only
*** adjusted Rank and DCE coefficients = DCE coefficient / Death dummy

Figure 1: Comparison of coefficients for TTO, rank and DCE model

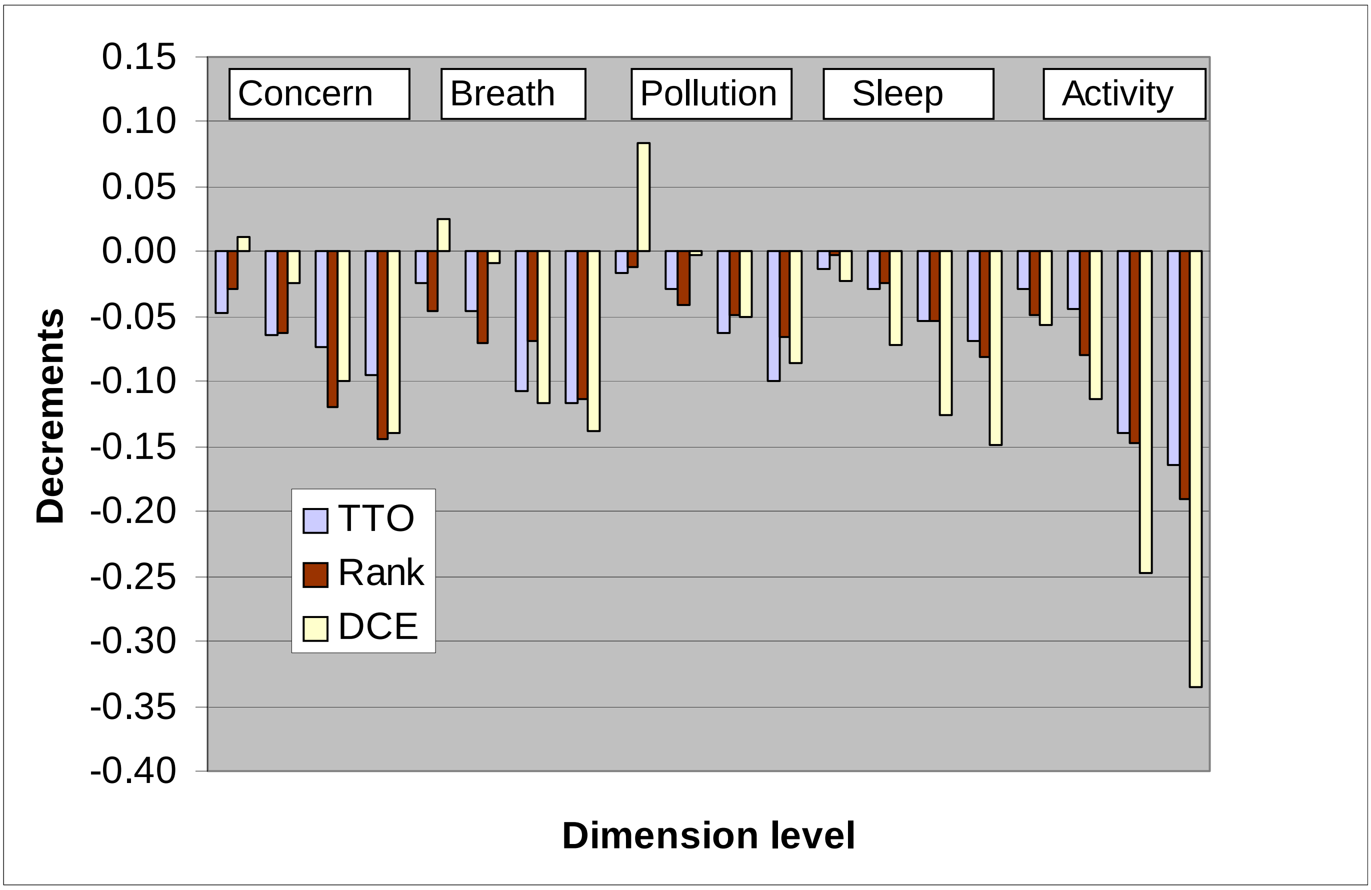

Figure 2: Health state value predictions for TTO, Rank and DCE models