A framework for estimating health state utility values within a discrete choice experiment: Modelling risky choices

Angela Robinson[1], Anne Spencer[2] & Peter Moffatt[1]

[1.] University of East Anglia

[2.] University of Exeter Medical School.

Corresponding author:
Angela Robinson
Norwich Medical School
University of East Anglia
Norwich
Norfolk
NR4 7TJ
e mail: angela.robinson@uea.ac.uk

Word count : 4781 (excluding the appendix)

Abstract

Background There has been recent interest in using the DCE method to derive health state utilities for use in QALY calculations but challenges remain. **Objectives** We set out to develop a risk-based DCE approach to derive utility values for health states that allowed; a) utility values to be anchored directly to normal health and death and b) worse than dead health states to be assessed in the same manner as better than dead states. Further, we set out to estimate alternative models of risky choice within a DCE model. **Method** A survey was designed that incorporated a risk-based DCE and a 'modified' SG. Health state utility values were elicited for three EQ-5D health states assuming 'standard' EU preferences. The DCE model was then generalised to allow for Rank Dependent Expected Utility (RDU) preferences thereby allowing for probability weighting. A convenience sample of 60 students was recruited and data collected in small groups. **Results** Under the assumption of 'standard' EU preferences, the utility values derived within the DCE corresponded fairly closely to the mean results from the modified SG. Under the assumption of RDU preferences, the utility values estimated are somewhat lower than under the assumption of standard EU, suggesting that the latter may be biased upwards. **Conclusion** Applying the correct model of risky choice is important whether a modified SG or a risk-based DCE is deployed. It is, however, possible to estimate a probability weighting function within a DCE and estimate 'unbiased' utility values directly which is not possible within a modified SG. We conclude by setting out the relative strengths and weaknesses of the two approaches in this context.

INTRODUCTION

Discrete choice experiments (DCEs) allow a number of characteristics to be traded off against one another and are becoming increasingly popular in health economics. Although the origins of the DCE approach lie in marketing, they were later applied to the valuation of aspects of health care not easily captured using conventional Quality of life measures, such as the type of services received[1,2]. There has, however, been recent interest in using the DCE method to derive health state utilities for use in QALY calculations and there remains uncertainty about how that is best done [3]. This paper adds to the small literature on how utility values for health may be derived directly within a DCE.

The motivation for the use of DCEs to elicit utility values has been linked to well-known problems that exist with more traditional value elicitation techniques such as Standard Gamble (SG) and Time Trade Off (TTO) [3-7]. For example, the SG has been criticized on the basis that Expected Utility (EU) is not a descriptively valid theory of decision making under risk [7]. Furthermore techniques (such as SG and TTO) that set out to elicit an individual's point of indifference are regarded as more cognitively demanding than those involving pair-wise choices [5,6].

There are a number of methodological challenges, however, to be addressed in deriving health state utility values in particular within a DCE. In more traditional applications of DCEs attributes can reasonably be assumed to contribute to utility in an additive way, albeit with the possibility of interaction effects between attributes. The inclusion of attributes such as risk [7] in a DCE, presents different challenges than have traditionally arisen in terms of the appropriate functional form of the model. Risk has to enter the model in a multiplicative, rather than additive, manner as there is no meaningful preference over 'risk', independent of what that risk is of.

Estimating utility values for health states within a DCE requires health states to be anchored to normal health (generally assigned a value of '1') and dead (generally assigned a value of 'zero'). There are DCE studies that look at comparison of health states, without trying to link them to a normal health/dead scale[8], but the results cannot then be used as utility values and incorporated into QALY calculations. Other

researchers have tried to link DCE valuations to normal health and death by asking one or more SG or TTO questions alongside the DCE [3]. If DCE is being applied in order to overcome perceived problems with SG and TTO, then using those methods to anchor valuations is clearly problematic.

Recent studies have used a 'TTO-like' format and have linked health states to normal health and dead within a DCE by either including 'dead' as a state or survival duration as an attribute[9,10]. The approach, termed DCE$_{TTO}$, includes health states and survival duration (not including zero) as attributes, but does not include immediate death as an option. Bansback and colleagues use this approach to value a range of EQ-5D states[9]. In the DCE$_{TTO}$, normal health is set at one and values worse than dead can be inferred indirectly at a sample level, by using the coefficients on attribute levels as incremental reductions in quality of life. Thus, setting normal health equal to 1 and subtracting incremental decreases in quality of life associated with attribute levels, there will come a point when the values lie below zero. Values inferred less than zero are then taken to signify that the state is worse than dead. One possible drawback with this approach is that respondents are never asked to consider death directly.

Another way in which utility values may be linked directly to death within a DCE is to include some *risk* of immediate death as an option. One advantage of using risky choices is that there is a body of research looking at decision making under risk[11,12] and there has been some success in adjusting for biases in risky choices[13,14]. Whilst traditional SG approaches generally involve the certainty of the 'target' health state, we follow Carthy et al[15] in using the term 'modified' SG here to denote that risk appears in both options (the approach is referred to elsewhere as the 'lottery equivalent' approach[16,17]).

If we are to use risk-based DCE, it is important to consider how the theory of random utility might be adapted to incorporate recent advances in decision-making under risk. The random utility theory of McFadden underpins the analysis of DCEs [18]. It models decision making as a stochastic process around expected utility. There are, however, a number of other models of decision making under risk. For example, Rank Dependent Expected Utility (RDU) assumes that people over- or under-weight probability and so

it incorporates a probability weighting function in its specification of decision making [19,20]. Cumulative Prospect Theory subsumes RDU as it incorporates a probability weighting function and also allows for people to experience greater changes in utility from losses compared to gains[21].

It has been argued previously that random utility theory can incorporate a broad range of preference models that can be estimated using discrete choice experiments[22]. One important paper considered the use of rank dependent expected utility functions in DCE in looking at treatments and side-effects of Crohn's disease[7]. The authors found evidence of non-linearity in how risks were perceived and derived lower utility values under assumptions of RDU than 'standard' EU (henceforth we use simply 'EU' to denote 'standard EU'). This led them to argue that traditional SG methods assuming EU are biased but stressed that it was difficult to do a direct comparison as the nature of the risks included in their DCE was very different to those commonly used in SG[7].

One criticism of traditional value elicitation techniques, such as SG and TTO is that the procedures for valuing states worse than dead involve a fundamental departure from those used to value better than dead states. Given the large body of evidence showing that responses can be affected by descriptive and procedural invariance[23] we argued previously that such evidence must call into question the validity of aggregating better than and worse than dead scores generated by two different procedures[24]. It can be shown, however, that a technique which presents respondents with choices over two risky treatments allows states worse than dead to be valued in the same manner as better than dead states. One such method is the 'modified' SG which also avoids any 'certainty effect' bias uncovered elsewhere [25-27] as it incorporates risk on both sides. This approach has been used successfully previously[15-17,28].

We set out here to develop a method whereby utility values may be derived within a DCE that anchors them directly to normal health and death. Further, we set out to allow worse than dead health states to be derived in the same manner as better than dead states. Whilst the nature of the risk attribute used in a previous risk-based DCE study was such that direct comparisons with SG were problematic[7], we set out here to make the methods as comparable as possible.

The aims of this research are therefore:

1) To develop a method for eliciting values for health states, anchored to normal health and dead, within a risk-based DCE.

2) To develop a framework in which values for 'better than dead' and 'worse than dead' health states can be elicited in the same manner.

3) To compare EU and RDU models of risky choice within a DCE.

4) To compare the results of the DCE model(s) with the modified SG.

2. METHODS

2.1 Overview of the survey

There were 60 participants recruited from the population of second and third year students studying Economics or Geography at the Universities of London (Queen Mary) and Exeter in 2011/12. Data were collected by means of small groups comprising on average between 8 and 9 participants which were convened by two authors (AS and AR). Respondents were invited to take part either through e-mail (at Queen Mary) or through the experimental laboratory (FEELE at Exeter University) and were each paid £10. The study was granted local ethical approval in both institutions.

The groups began with a brief introduction to the aims of the study and the the EQ-5D health states (21121, 22222 and 22323). Respondents were first asked to rank the health states along with normal health (11111) and 'immediate death'. This was followed by DCE questions (15) and modified SG questions (3). The order in which the DCE and modified SG questions appeared was randomised. Finally respondents answered a series of 4 questions designed to elicit risk attitudes using money lotteries.

2.2 The DCE questions

In the DCE exercise, respondents were presented with two risky treatments, labeled A and B. All risky treatments involved some chance p of an outcome (21121, 22222, 22323, or immediate death) and an associated chance, 1-p, of normal health (11111). A typical question is shown in Figure 1 and used graphical displays to illustrate risk information. In this case, Treatment A offers a 10% chance of normal health and a

corresponding 90% chance of health state 21121. Treatment B offers a 99% chance of normal health and 1% chance of immediate death. We simplify this notation henceforth as Treatment A offers a 90% chance of 21121 and Treatment B offers a 1% chance of death. It is important to remember, however, that there is always an associated chance of normal health as normal health appeared in *all* treatments. In this example and at the point of indifference, the calculation (under EU) is: $0.9 (U_{21121}) + 0.1(U_{11111}) = 0.01 (U_{dead}) + 0.99(U_{11111})$ and assigning values of 0 and 1 to dead and normal health respectively and rearranging gives: $0.9 (U_{21121}) = 0.89$, so $U_{21121} = 0.988$

Respondents were asked to suppose that they had some condition and they were faced with two different treatments for that condition. They were asked to tick one of three possible responses, namely: prefer A; equally preferable, prefer B. We elected to include the 'indifference' option in the choice data as we wanted to maximize the similarities across the DCE and modified SG approaches. This avoids 'forcing' a preference which may not always be appropriate[29].

Whilst we have used the EQ-5D descriptive system for convenience, it is important to stress here that we are *not* setting out here to derive a set of weights for that system. As developing a methodology is our aim here, we opted for a very simple design involving only two attributes- outcome and risk. The DCE questions varied on one or more of the two attributes shown below:

- The outcome (health states 21121, 22222, 22323 or immediate death) coloured yellow, green, grey and blue respectively.
- The probability of that outcome (1%, 5%, 10%, 20%, 30%, 40%, 50%, 70%, 90%).

The attributes and levels set in this study produced a total of 630 different combinations (as there are $9 \times 1 \times 4 = 36$ scenarios and the number of ways of choosing r=2 scenarios at random from n=36 is $(n)!/(n-r)!r! = 36.35/2 = 630$). We chose to include all non-dominated combinations in this exploratory study. This was due to the uncertainty surrounding the optimal design for multiplicative models [30] and to avoid

limiting the scope of the analysis unnecessarily by using a fractional factorial design[29].

There were three types of dominance that arose in the study due to the levels of risk, health states or both. Choices could be 'risk-dominated' in that they involved the same health state but a different level of risk attached to that state. We elected to ask all participants (a different) one of the 144 'risk-dominated' questions contained in the full factorial, as a simple test of consistency, and so included 60 risk-dominated choices. As there is a 'logical' ordering of health states in that $21121 \succ 22222 \succ 22323$, choices could be 'state-dominated' in that they involved the same level of risk but a better/worse health state. The full factorial contained a total of 27 such choices which we randomly allocated. Finally, choices could be 'risk/state dominated' in that they involved a lower risk of a less severe state. For example, suppose that Treatment A offered a 30% chance of health state 21121 (and associated 70% chance of normal health) and Treatment B offered a 40% chance of 22222 (and associated 60% chance of normal health). Treatment A clearly dominates Treatment B, as is offers a lower risk of less severe illness. The full factorial contained a total of 108 such comparisons which we retained and which were randomly allocated. In total our design included 546 choices (630-144=546): 351 non-dominated and 195 dominated choices that we elected to retain (60+27+108=195).

Respondents were presented with a set of 15 DCE questions. The first question was one drawn randomly from the 144 'risk dominated' comparisons described above. Each respondent was presented with a further 8 questions drawn randomly from the full factorial design. In addition, respondents were presented with a common set of 6 questions that were interspersed with those randomly allocated. The 6 'common' questions were a series that set out to allow the utility value of one health state-22222- to be determined at the level of the individual respondent. These questions are not the focus of the current paper.

2.3 The modified SG

In the modified SG part of the questionnaire, the framing of the question was designed to closely resemble the pair-wise choices that appeared in the DCE. Rather than having the risks associated with both treatments fixed in advance and being

asked to choose between treatments, only one treatment was fixed in the modified SG. Respondents were presented with a fixed risk of the health state under Treatment A, and then asked to 'set' that risk of death in Treatment B that made them indifferent between the two treatments. Participants were asked three modified SG questions. For health states 21121 and 22222, Treatment A involved a 90% risk of that state. For health state 22323, Treatment A involved a 20% risk of that health state, to allow for potentially lower values. The modified SG questions were asked in a fixed order 21121, 22222 and then 22323. Groups were randomized to see DCE or modified SG first.

Utility values are then estimated directly from the modified SG in exactly the same way as set out above. Considering the choice set out in Figure 2, suppose the respondent sets the indifference probability of dead at 0.20, then under EU :

$0.90 (U_{22222}) + 0.10(U_{11111}) = 0.20 (U_{dead}) + 0.80(U_{11111})$ and assigning values of 1 and 0 to full health and dead respectively gives: $(U_{22222}) = 0.78$.

The format of both the modified SG and DCE questions allow worse than dead states to be valued in exactly the same manner as better than dead states. For example, suppose the modified SG question involved a 20% risk of EQ-5D health state 22323 under Treatment A, and the respondent set the risk of death under Treatment B at 40%. Then $0.2 (U_{22323}) + 0.8(U_{11111}) = 0.40 (U_{dead}) + 0.60(U_{11111})$ and assigning value of 1 and 0 to normal health and dead respectively gives: $(U_{22323})= -1$. In effect, health state 22323 is worse than dead, provided that respondents prefer to take a higher risk of death to avoid the risk of health state 22323.

In the final part of the questionnaire, four questions were used to elicit participants' risk attitudes for monetary lotteries, using the mid-weight method proposed by Kuilen and Wakker[31]. These questions are not the focus of the current paper, but details are available from the authors on request.

2.4 Modelling the DCE Choices

The estimation strategy is discussed in detail in the appendix: we provide a non-technical summary here. Recall that respondents were presented with two risky treatments, labeled A and B. All risky treatments involved some chance p of an

outcome (21121, 22222, 22323, or immediate death) and an associated chance, 1-p, of normal health (11111). They were asked to tick one of three possible responses, namely: prefer A; equally preferable, prefer B. Briefly, we used an ordered probit model to assess the likelihood of each response based on the difference in the expected utility of treatments A and B (the EU model). We consider both EU and RDU preferences, the latter allowing for non-linear weighting of probabilities. In the RDU model, rather than calculate the expected utility of treatments A and B, we first apply a transformation mechanism to the probabilities in the form of a power weighting function $\pi(r) = r^\gamma$ where r is the probability of the good outcome (in this particular case equal to 1-p). This weighting function implies that the probability of the good outcome, when strictly between 0 and 1, is either always over-weighted (if $\gamma<1$) or always underweighted (if $\gamma>1$). A strictly convex weighting function has been estimated previously [32].

The parameters of central interest are the utilities of the three health states 21121, 22222 and 22323, denoted by u_1, u_2 and u_3 respectively.

The parameter κ is known as the "cut-point", and indicates the distance from perfect indifference a subject must be to indicate a clear preference between the two alternatives (closer to indifference, 'equally preferable' is reported).

There is a further way in which we generalize the econometric model. The model contains a parameter σ which represents computational error by the respondent in computing the valuation difference for a given problem. We first estimate the models assuming that this parameter takes the same value for all problems. We then relax this assumption by allowing σ to differ between different problem types as it seems plausible that errors are more likely with some problem types than others. The inclusion of 'dominated' options may influence this as may the inclusion of choices involving 'immediate death' which may well be perceived very differently than those involving two health states. Therefore, we allowed three different error variance parameters: $\sigma_{dominance}$ for "dominance" problems; σ_{death} for problems involving 'immediate death' as one of the outcomes and $\sigma_{standard}$ for "standard" problems (i.e. non dominance problems not involving immediate death).

3. RESULTS

3.1 Estimation of utility values from the modified SG questions

Recall that in the modified SG questions respondents were presented with a fixed risk of a health state under Treatment A, but then asked to 'set' that risk of death in Treatment B that made them indifferent between the two treatments. Using the EU calculations set out in the methods section above, the utility value of the health states can be calculated for each individual. Table 1 presents mean and median utility values for the 3 health states from the modified SG assuming standard EU preferences as is traditional. Of course, if the EU model is not correct and respondents were under-weighting probabilities, then the modified SG results will over-estimate the utility values. We return to this in the discussion.

3.2 Estimation of utility values within the DCE model

The results of the DCE models are presented in Table 2 which shows the results under both EU and RDU, with both the fixed error and varying error specifications.

The parameter estimates of primary interest are the three utilities. Firstly note that the utility estimates obtained under the assumption of EU (σ fixed) are fairly close to the corresponding mean utility values from the modified SG presented in Table 1, indicating a degree of consistency across methods. Note also that utility values estimated under the assumption of RDU are somewhat lower than under the assumption of standard EU, indicating that the latter may be biased upwards. Recall that the power weighting function implies that the probability of the good outcome, when strictly between 0 and 1, is either always over-weighted (if $\gamma < 1$) or always underweighted (if $\gamma > 1$). The latter turns out to be the case for this data set, a situation sometimes referred to as strictly convex weighting. The estimate the power-weighting parameter γ (taken from the final column of results) is 3.747 and the implied weighting function is shown in Figure 3, together with the 45^0-line which is implied under EU. It is clear that the probability of normal health is seriously under-weighted, particularly when the true probability is small.

It must be recognised that the weighting function depicted in Figure 3 does not capture the "inverse-S" shape that has become standard in the modelling of choices over money gambles. However, when the inverse-S function of Tversky and Kahneman [21] is applied to this problem, the fixed point (where the weighting function crosses the 45^0-line) is found to be very close to the origin, and consequently the function is not dissimilar to the power function with $\gamma > 1$. Systematic under-weighting of probabilities has been found in different contexts previously [32,33]. It is important to note that, if utilities derived under standard EU assumptions are biased upwards, that will hold for the modified SG results also.

The third and fourth columns of Table 2 present the results of the models in which the variance parameter is assumed to vary between problem types. Firstly, note that on the evidence of Akaike's Information Criterion (AIC) which is reported in the final row, this assumption results in a considerable improvement in statistical fit even allowing for the presence of additional parameters. Unsurprisingly perhaps, dominance problems have the smallest estimated variance. Problems involving death have the highest estimated variance and we return to this in the discussion.

We conclude by reporting that only 2 (of 60) respondents failed any of the dominance tests, although it is not too surprising that a sample of students (many of whom had studied economics) would pass such tests.

4. DISCUSSION

We report the results of an exploratory study that set out to develop a risk-based DCE to derive utility values for health states and to compare the results with those from a modified SG. Both methods allow health states to be anchored to normal health and death, allowing utility values to be derived directly using either method. Further, both the modified SG and risk-based DCE deployed here allowed worse than dead states to be valued in the same manner as better than dead states. Our results show a broad correspondence between the results from DCE model and the mean modified SG results, particularly under the assumption of EU preferences. The results are very

similar indeed for two of the health states (21121 and 22222) whilst the DCE model results are higher than modified SG for 22323. Of course, demonstrating correspondence between the results does not, in itself, allow anything to be concluded about the relative merits of the different approaches.

It is widely considered that SG values ought to be corrected for probability-weighting if utilities are not to be biased upwards[14]. It is clear from our results that the same is true for utilities derived within a DCE and we have demonstrated how utility values may be adjusted for probability weighting directly in a DCE. Our findings are consistent with those of other studies that have shown consistent underweighting of probability both in the area of health[33] and other risky choices[31].

Whilst it is possible to adjust modified SG values to allow for probability weighting, this would rely on using a probability-weighting function derived elsewhere. This obviously begs the question of where such a probability-weighting function would come from. Although not a prominent part of this paper, we did explore the use of risk attitude money lottery questions that would potentially allow a within-sample probability weighting factor to be derived and used to adjust SG valuations[31]. An obvious methodological issue there would be whether risk attitudes in the domain of money lotteries would necessarily be the same as those in health[34]. In addition, the feasibility of asking the required number of money lottery questions alongside the SG method would have to be established. It does appear, therefore, that a DCE incorporating an appropriately specified model of risky choice offers the more promising means of deriving 'unbiased' utility values in this context.

We found an improvement in statistical fit of the DCE model in allowing the error variance to differ between different problem types. Problems involving death were found to have the highest estimated error variance, which may indicate that those questions were more difficult for respondents. The conceptual problems of including 'dead' as a health state in a DCE have been outlined previously[30], but utility values have to be anchored to 1 and 0 somehow. Whilst we have demonstrated a framework for anchoring utility values directly to normal health and dead – and of valuing better

and worse than dead states using exactly the same procedure – it is clear that there remain challenges in incorporating death directly into a DCE.

We conclude by making some more general points about the assessment of the relative merits of DCE and more traditional methods such as SG and TTO. Methods such as SG and TTO are traditionally thought of as 'matching' techniques – whereby the task is to 'set' the level of risk/duration that makes the respondent indifferent between two options. There is a literature on the fact that 'matching' and 'choice' tasks maybe tapping into different cognitive processes and, hence, the results are likely to differ across methods. One criticism of 'matching' tasks is that, in asking respondents to 'match' on any single dimension encourages respondents to attach undue weight to that specific dimension while neglecting other factors that they would otherwise wish to be taken into consideration[23].

Whilst the modified SG that respondents completed here was an actual 'matching' task (in that we asked respondents to directly 'set' the probability of death in Treatment B to make them indifferent between A and B), it is important to acknowledge that most SG and TTO elicitation techniques actually present respondents with a series of pair-wise choices. Holding the format of the questions the same, the only difference between SG (or TTO) and a DCE is that in the former the choices are generally generated by an interactive process that tries to 'hone in' on that respondents point of indifference. When considered in this way, the SG (and TTO) could be seen as more 'efficient' techniques at arriving at utility values than DCE. On the other hand, a possible drawback of any interactive approach is that 'starting point' biases and anchoring effects may be introduced that are avoided in the DCE.

It is not possible to choose between the methods on the basis of the current study. As there was a logical ordering of the three EQ 5D states used here, the ranking data offers little by way of 'validating' the findings-although this could offer a useful check in future studies. We cannot offer an assessment of whether the modified SG or DCE incorporating risky choices is 'superior', but we can go some way towards setting out the potential strengths and weaknesses in this context. Setting out valid arguments is a small, but important, step as we believe certain arguments that have

been put forward previously are not valid. For example, arguing that methods such as SG (and TTO) that set out to 'hone in' on a point of indifference- i.e. where it is *hardest* to choose between options -are 'harder' for respondents to do than DCE [6] is somewhat spurious. After valid arguments have been set out, the choice between methods may ultimately depend on the relative weights attached to the various criteria and how they are traded-off against one another. For example, if it is of primary importance to investigate utility values at the level of the individual, then SG may be favoured over DCE. On the other hand, having a relatively simple means of adjusting aggregate values to allow for probability may be considered of more importance, in which case the DCE will be favoured. We present a summary of the main strengths and weaknesses of the two methods of deriving utility values in Table 3 below. Additional considerations would be the ability to value 'process' factors alongside health outcomes (which would favour DCE [35]) and resource efficiency in terms of the sample size requirements, but these issues are beyond the scope of this paper.

An obvious limitation of the study is that we used a convenience sample of students and it remains to be seen how members of the public would cope with the exercises. We only derived valuations for three EQ 5D health states, but the method could be expanded in order to allow for the valuation of attribute levels and a 'tariff' to be estimated.

We believe the body of work currently being undertaken to estimate utility values within a DCE is important and that the method does have the potential to replace traditional methods such as SG and TTO. We believe we have gone some way towards setting out the methodological issues that remain and in clarifying the strengths and weaknesses of DCE compared to more traditional methods.

Bibliography

1.   Damman OC, Spreeuwenberg P, Rademakers J, Hendriks M. Creating Compact Comparative Health Care Information: What Are the Key Quality Attributes to Present for Cataract and Total Hip or Knee Replacement Surgery? *Med. Decis. Mak.* Mar-Apr 2012;32:287-300.
2.   Scott A, Watson MS, Ross S. Eliciting preferences of the community for out of hours care provided by general practitioners: a stated preference discrete choice experiment. *Soc. Sci. Med.* Feb 2003;56:803-814.
3.   Stolk EA, Oppe M, Scalone L, Krabbe PFM. Discrete Choice Modeling for the Quantification of Health States: The Case of the EQ-5D. *Value Health.* Dec 2010;13:1005-1013.
4.   Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. *Journal of health economics.* Jan 2011;31:306-318.
5.   Norman R, Viney R, Brazier J, et al. Valuing SF-6D Health States Using a Discrete Choice Experiment. *Medical decision making : an international journal of the Society for Medical Decision Making.* Sep 11 2013.
6.   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):15(13)
7.   Van Houtven G, Johnson FR, Kilambi V, Hauber AB. Eliciting Benefit-Risk Preferences and Probability-Weighted Utility Using Choice-Format Conjoint Analysis. *Med. Decis. Mak.* May-Jun 2011;31:469-480.
8.   Hakim Z, Pathak DS. Modelling the EuroQol data: A comparison of discrete choice conjoint and conditional preference modelling. *Health Econ.* Mar 1999;8:103-116.
9.   Bansback N, Tsuchiya A, Brazier J, Anis A. Canadian Valuation of EQ-5D Health States: Preliminary Value Set and Considerations for Future Valuation Studies. *PLoS One.* 2012 (Epub 2012 Feb 2012;7:e31115.
10.   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. *The European journal of health economics : HEPAC : health economics in prevention and care.* Oct 2011;13:575-587.
11.   Karni E. A theory of medical decision making under uncertainty. *Journal of Risk and Uncertainty.* Aug 2009;39:1-16.
12.   Machina MJ. Choice under uncertainty - problems solved and unsolved - responses. *J. Econ. Perspect.* Spr 1988;2:181-183.
13.   Abellan-Perpinan JM, Bleichrodt H, Pinto-Prades JL. The predictive validity of prospect theory versus expected utility in health utility measurement. *J. Health Econ.* Dec 2009;28:1039-1047.

14. Doctor JN, Bleichrodt H, Lin HJ. Health Utility Bias: A Systematic Review and Meta-Analytic Evaluation. *Med. Decis. Mak.* Jan-Feb 2010;30:58-67.
15. Carthy T, Chilton S, Covey D, et al. On the contingent valuation of safety and the safety of contingent valuation: Part 2 - The CV/SG "chained" approach. *Journal of Risk and Uncertainty.* 1998;17:187-213.
16. Law AV, Pathak DS, McCord MR. Health status utility assessment by standard gamble: a comparison of the probability equivalence and the lottery equivalence approaches. *Pharmaceutical research.* Jan 1998;15:105-109.
17. McCord M, Deneufville R. Lottery equivalents -reduction of the certainty effect problem in utility-assessment. *Management Science.* Jan 1986;32:56-60.
18. McFadden D. Conditional Logit Analysis of Qualitative Choice Behavior. *in P. Zarembka (ed.), Frontiers of Econometrics, Academic Press.* 1974.
19. Quiggin J. *Generalized Expected Utilty Theory. The Rank-Dependent Model.* Boston: Kluwer academic Publishers; 1993.
20. Schmeidler D. Subjective probability and expected utility without additivity. . *Econometrica.* 1989;57.(3):571-587.
21. Tversky A, Kahneman D. Advances in prospect theory - cumulative representation of uncertainty. *Journal of Risk and Uncertainty.* Oct 1992;5:297-323.
22. de Palma A, Ben-Akiva M, Brownstone D, et al. Risk, uncertainty and discrete choice models. *Mark. Lett.* Dec 2008;19:269-285.
23. Tversky A, Sattath S, Slovic P. Contingent weighing in judgment and choice. *Psychol. Rev.* Jul 1988;95:371-384.
24. Robinson A, Spencer A. Exploring challenges to TTO utilities: valuing states worse than dead. *Health Econ.* 2006;15:393-402.
25. Hershey JC, Schoemaker PJH. Probability versus certainty equilvalence methods in utility measurement - are they equivalent. *Management Science.* 1985;31:1213-1231.
26. Van Osch SMC, Stiggelbout AM. The construction of standard gamble utilities. *Health Econ.* Jan 2008;17:31-40.
27. van Osch SMC, van den Hout WB, Stiggelbout AM. Exploring the reference point in prospect theory: Gambles for length of life. *Med. Decis. Mak.* Jul-Aug 2006;26:338-346.
28. Bleichrodt H, Abellan-Perpinan JM, Pinto-Prades JL, Mendez-Martinez I. Resolving inconsistencies in utility measurement under risk: Tests of generalizations of expected utility. *Management Science.* Mar 2007;53:469-482.
29. Viney R, Lancsar E, Louviere J. Discrete choice experiments to measure consumer preferences for health and healthcare. *Expert review of pharmacoeconomics & outcomes research.* Aug 2002;2:319-326.
30. Flynn TN. Using Conjoint Analysis and Choice Experiments to Estimate QALY Values Issues to Consider. *Pharmacoeconomics.* 2010;28:711-722.
31. van de Kuilen G, Wakker PP. The Midweight Method to Measure Attitudes Toward Risk and Ambiguity. *Management Science.* Mar 2011;57:582-598.
32. Goeree JK, Holt CA, Palfrey TR. Quantal response equilibrium and overbidding in private-value auctions. *Journal of Economic Theory.* May 2002;104:247-272.

33. Bleichrodt H. Probability weighting in choice under risk: An empirical test. *Journal of Risk and Uncertainty.* Sep 2001;23:185-198.

34. Prosser LA, Wittenberg E. Do risk attitudes differ across domains and respondent types? *Med. Decis. Mak.* May-Jun 2007;27:281-287.

35. Ryan M. Discrete choice experiments in health care: NICE should consider using them for patient centred evaluations of technologies. *British Medical Journal* 2004 328 (7436):360-361.

Table 1 Mean, median and standard deviation of utility values from modified SG

EQ -5D state	Mean	Median	SD
21121	0.899	0.944	0.130
22222	0.816	0.899	0.168
22323	0.214	0.500	0.716

Table 2 Estimates of coefficients (asy. st. errors) from DCE models

	EU (σ fixed)	RD(σ fixed)	EU (σ varying)	RD(σ varying)
u_1 (21121)	0.907(0.030)	0.817(0.041)	0.882(0.018)	0.799(0.031)
u_2 (22222)	0.789(0.015)	0.592(0.043)	0.786(0.015)	0.621(0.038)
u_3 (22323)	0.284(0.050)	0.169(0.049)	0.336(0.052)	0.245(0.051)
σ	0.154(0.010)	0.276(0.022)		
$σ_{standard}$			0.106(0.019)	0.193(0.036)
$σ_{dominance}$			0.042(0.008)	0.107(0.021)
$σ_{death}$			0.163(0.011)	0.289(0.025)
γ		4.253(0.829)		3.747(0.689)
κ	0.031(0.003)	0.059(0.007)	0.025(0.003)	0.052(0.007)
N	900	900	900	900
K	5	6	7	8
LogL	-613.53	-582.88	-595.07	-568.20
AIC (2k-2LogL)	1237.06	1177.76	1204.14	1152.4

Table 3 Summary of the relative strengths and weaknesses of DCE and modified SG

Advantage	Modified SG	Risk-based DCE
Allows health states to be anchored directly to normal health and death	✓	✓
Allows WTD and BTD states to be assessed on the same scale	✓	✓
Individual-level utility values generally derived*	✓	✗
Relatively few questions need be asked	✓	✗
Weighting function internal to model	✗	✓
Less susceptible to starting-point bias	✗	✓

*We recognize that it is possible to derive individual-level valuations within a DCE by asking respondents to answer a large number of questions-but that is not what DCEs *typically* do.

Figure 1 An example of a DCE pair-wise comparison

Treatment A

10 in 100 (10%) chance of pink state
  No problems walking about ☐
  No problems with self-care ☐
  No problems with performing usual activities ☐
  No pain or discomfort ☐
  Not anxious or depressed ☐
              And
90 in 100 (90%) chance of yellow state
  Some problems walking about ▨
  No problems with self-care ☐
  No problems with performing usual activities ☐
  Moderate pain or discomfort ▨
  Not anxious or depressed ☐

Treatment B

99 in 100 (99%) chance of pink state
  No problems walking about ☐
  No problems with self-care ☐
  No problems with performing usual activities ☐
  No pain or discomfort ☐
  Not anxious or depressed ☐
              and
1 in 100 (1%) chance of death

☐
Choose to have

☐
Both options are equally preferred

☐
Choose to have

Figure 2 Modified SG question

Treatment A

10 in 100 (10%) chance of pink state
- No problems walking about ☐
- No problems with self-care ☐
- No problems with performing usual activities ☐
- No pain or discomfort ☐
- Not anxious or depressed ☐

and

90 in 100 (90%) chance of green state
- Some problems walking about ▢
- Some problems washing and dressing self ▢
- Some problems with performing usual activities ▢
- Moderate pain or discomfort ▢
- Moderately anxious or depressed ▢

Treatment B

[   ] chance of pink state
- No problems walking about ☐
- No problems with self-care ☐
- No problems with performing usual activities ☐
- No pain or discomfort ☐
- Not anxious or depressed ☐

and

[   ] chance of death

[   ] chance of death

Both options are equally preferred
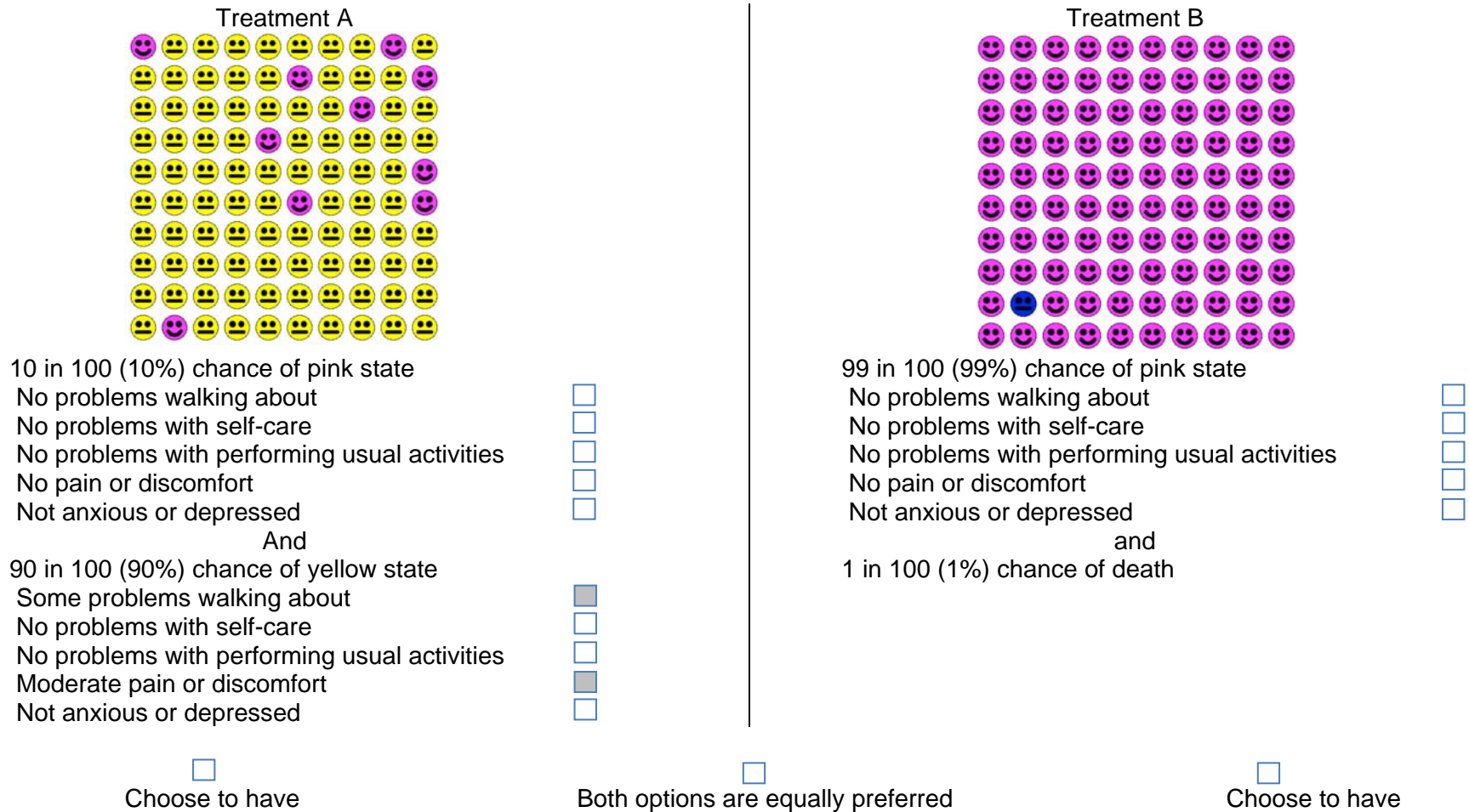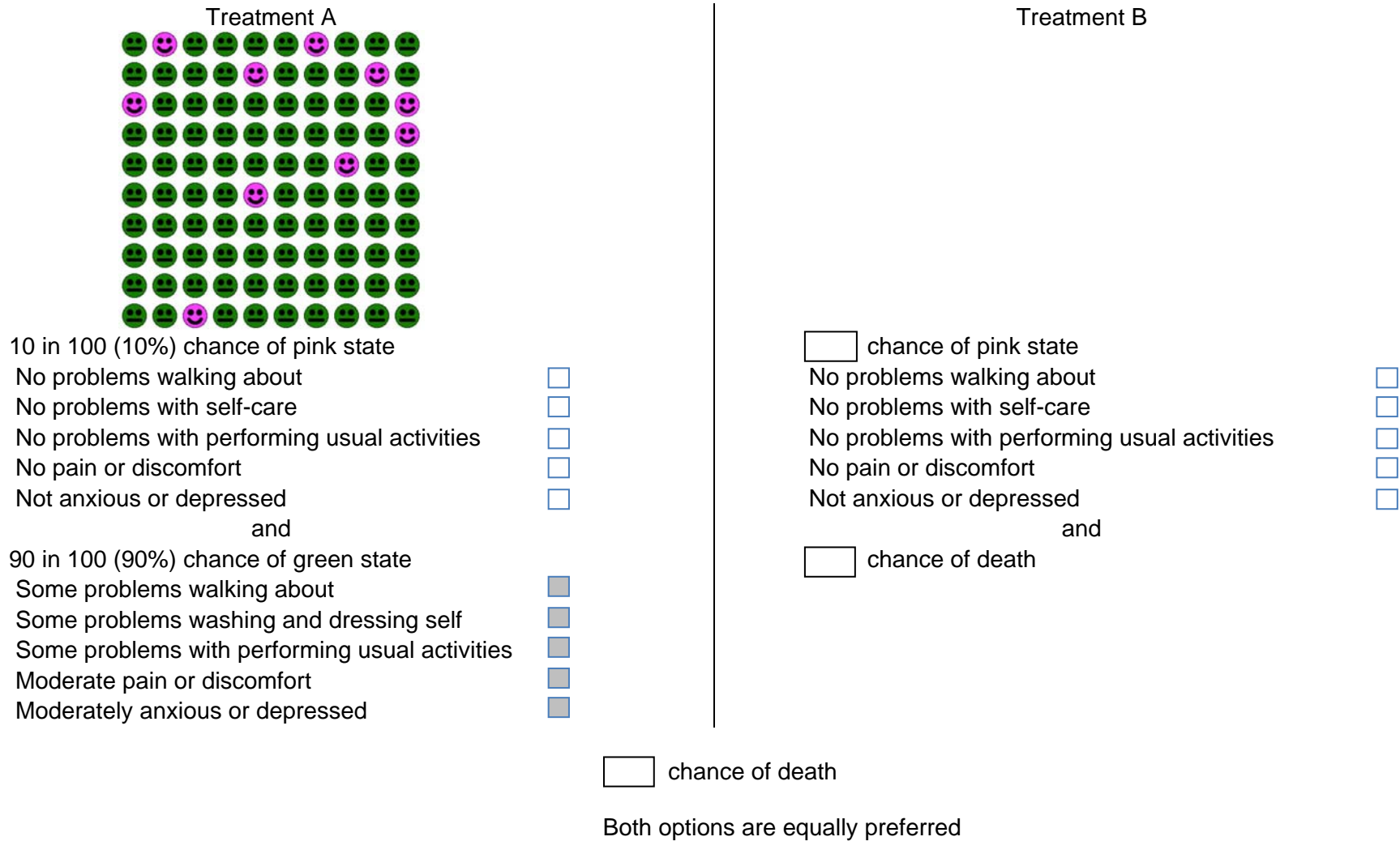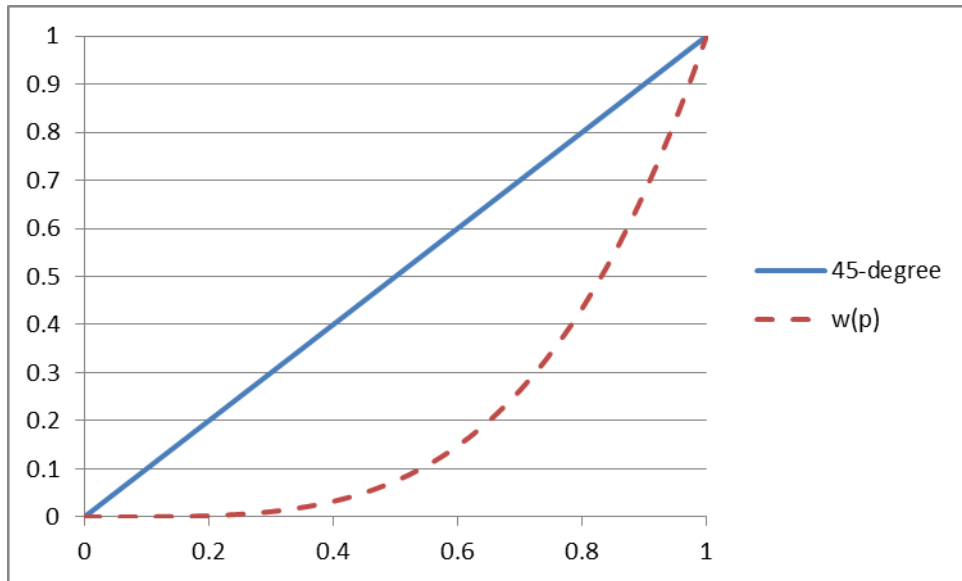
Figure 3: the probability weighting function

Appendix: Modelling the DCE Choices

For this section, we introduce more formal notation. Let X_j denote health-state j. The models developed in this section make the standard, simplifying assumption that all individuals have the same utility value for a given health state. Recall that we are commencing from the 'anchors' of the utilities of normal health (X_0) and dead (X_4) being 1 and 0 respectively. There are three other health-states, X_1, X_2 and X_3, with utilities u_1, u_2 and u_3 respectively. The notation is provided in Table 4. The principal objective of the modeling is to obtain estimates of u_1, u_2 and u_3 i.e. to estimate utility values for health states anchored to normal health and death.

Consider choice problem i of the DCE. The choice is between two risky treatments A_i and B_i, defined as follows:

A_i: Probability $p_{a,i}$ of health state $X_{a,i}$; probability ($1- p_{a,i}$) of health state X_0.
B_i: Probability $p_{b,i}$ of health state $X_{b,i}$; probability ($1- p_{b,i}$) of health state X_0.

Under the assumption of EU, the individual computes valuations of A_i and B_i as follows:

$$EU(A_i) = p_{a,i}U(X_{a,i}) + (1 - p_{a,i})U(X_0)$$
$$EU(B_i) = p_{b,i}U(X_{b,i}) + (1 - p_{b,i})U(X_0) \qquad (1)$$
$$\nabla_i = EU(B_i) - EU(A_i)$$

Note that the symbol ∇_i is used to represent the difference in expected utilities. Let y_i denote the decision. Recall that there are three possible outcomes: prefer A ($y_i = 1$); A and B equally preferable ($y_i = 2$); prefer B ($y_i = 3$). We model this decision using a version of the ordered probit model developed by Aitchison and Silvey[36] defined as follows:

$$y_i = 1 \text{ if } \nabla_i + \varepsilon_i < \kappa$$
$$y_i = 2 \text{ if } -\kappa < \nabla_i + \varepsilon_i < \kappa$$
$$y_i = 3 \text{ if } \nabla_i + \varepsilon_i > \kappa$$

1

where $\varepsilon_i \sim N\left(0, \sigma^2\right)$ (2)

The parameter κ is known as the "cut-point", and indicates the distance from perfect indifference ($\Delta_i = 0$) within which "equally preferable" is reported. ε_i is a normally distributed random error term with standard deviation σ. As explained in the main text, we also consider a generalization of model (2) which allows sigma to take on a different value between three task types.

From (2), the probabilities of the three outcomes are derived as follows:

$$P\left(y_i = 1\right) = \Phi\left(\frac{-\kappa - \nabla_i}{\sigma}\right)$$

$$P\left(y_i = 2\right) = \Phi\left(\frac{\kappa - \nabla_i}{\sigma}\right) - \Phi\left(\frac{-\kappa - \nabla_i}{\sigma}\right) \qquad (3)$$

$$P\left(y_i = 3\right) = 1 - \Phi\left(\frac{\kappa - \nabla_i}{\sigma}\right)$$

where $\Phi(.)$ is the standard normal cumulative distribution function. From (3), the log-likelihood is constructed as follows:

$$LogL = \sum_i \left[\begin{array}{l} I\left(y_i = 1\right)\ln\Phi\left(\frac{-\kappa - \nabla_i}{\sigma}\right) + I\left(y_i = 2\right)\ln\left(\Phi\left(\frac{\kappa - \nabla_i}{\sigma}\right) - \Phi\left(\frac{-\kappa - \nabla_i}{\sigma}\right)\right) \\ + I\left(y_i = 3\right)\ln\left(1 - \Phi\left(\frac{\kappa - \nabla_i}{\sigma}\right)\right) \end{array} \right] \qquad (4)$$

The log-likelihood function (3) is programmed using the ML routine in STATA. The code is available from the authors on request. The parameters that are estimated are the utilities of the three health states other than normal health and death, i.e. u_1, u_2 and u_3, and also κ and σ.

As mentioned previously, we also consider a non-EU theory, in the form of RDU, which allows for non-linear weighting of probabilities. Here, we assume the straightforward power weighting function. If r is the probability of the good outcome (i.e. normal health in this case), then r is transformed according to:

$$\pi(r) = r^{\gamma} \qquad (5)$$

Note that the power weighting function (5) implies that the probability of the good outcome, when strictly between 0 and 1, is either always over-weighted (if $\gamma<1$) or always underweighted (if $\gamma>1$). Applying the weighting function (5), we derive the valuations of the two treatments:

$$V(A_i) = \left[1 - \pi\left(1 - p_{a,i}\right)\right] U\left(X_{a,i}\right) + \pi\left(1 - p_{a,i}\right) U(X_0)$$
$$V(B_i) = \left[1 - \pi\left(1 - p_{b,i}\right)\right] U\left(X_{b,i}\right) + \pi\left(1 - p_{b,i}\right) U(X_0) \qquad (6)$$
$$\nabla_i^{RD} = V(B_i) - V(A_i)$$

∇_i^{RD} is then used in place of ∇_i in the log-likelihood function (4), and of course there will be one additional parameter, γ, to be estimated.