

Department of Economics

Are Some Deaths Worse Than Others? Results from a Discrete Choice Experiment

Angela Robinson, Judith Covey, Anne Spencer and Graham Loomes

Working Paper No. 597

May 2007

ISSN 1473-0278



Queen Mary
University of London

Are some deaths worse than others? Results from a discrete choice experiment

Angela Robinson¹, Judith Covey², Anne Spencer³, Graham Loomes¹

¹ University of East Anglia

² University of Durham

³ Queen Mary, University of London

May 2007

ABSTRACT

Previous research has shown that people wish a premium to be placed on the prevention of certain types of deaths as they perceive those deaths as ‘worse’ than others. The research reported in this paper is an attempt to quantify such a ‘bad death’ premium via a discrete choice experiment (DCE). The four underlying attributes included were: the age of the victim, who was most to blame for the death, the severity of the victim’s pain and suffering in the period leading up to death, and the duration of the victim’s pain and suffering in the period leading up to death. In addition, a fifth attribute - number of deaths- was included in order to provide a *quantitative* scale against which to measure the “bad death premium”. The results show that each of the 4 underlying attributes did matter to respondents in determining whether deaths were worse than others, but also uncovered marked insensitivity to variations in the number of those deaths. The implication of our findings for the use of quantitative variables in DCEs is discussed.

Key words: discrete choice experiment, value of preventing a fatality, relative weights, insensitivity.

JEL classification: H5, I10.

1. Introduction

People regard some deaths as worse than others (Sunstein, 1997). For example, research has shown that a premium is placed on reducing the risks of, and saving lives from, certain types of deaths. This includes a higher premium for cancer deaths over deaths from heart disease, motor vehicle accidents, household fires or airplane accidents (Jones-Lee, Hammerton & Philips, 1985; Savage, 1993), deaths of cyclists over deaths of construction workers (Mendeloff & Kaplan, 1989) and deaths of 20-year olds over the deaths of 60-year olds (Cropper, Aydede & Portney, 1994). Likewise, studies have shown a higher premium for airplane deaths over deaths from heart attacks (Sunstein, 1997), deaths caused by exposure to industrial air pollution over deaths caused by smoking or automobile accidents (Subramanian & Cropper, 2000) and deaths in car accidents and fires in public places over deaths in domestic fires and railway accidents (Chilton, Covey, Hopkins et al., 2002). Other studies have shown a higher premium for deaths from lung disease due to industrial air pollution over deaths from liver disease due to contaminated drinking water (Hammit & Liu, 2004) and deaths in airplane journeys over deaths in taxi journeys (Carlsson, Johansson-Stenman & Martinsson, 2004).

These studies have suggested a number of explanations as to why people are more averse to certain types of deaths than others. The evidence suggests that they are concerned particularly about those deaths which affect them – or people close to them – personally (Mendeloff & Kaplan, 1989; Subramanian & Cropper, 2000; Chilton et al., 2002), kill young people (Cropper et al., 1994; Mendeloff & Kaplan, 1989), make people uneasy or feel “dread” when thinking about them (Savage, 1993; Carlsson et al., 2004; Chilton et al., 2002), and where the victims find it difficult to avoid the risk or are less to blame for their death (Subramanian & Cropper, 2000; Chilton et al., 2002).

Whilst this research gives us a sense of what makes specific types deaths worse than others, we are a long way from quantifying any “bad death premium” people may attach to the underlying features or attributes of these deaths (cf. Sunstein, 1997). Quantification

of this type would allow more generic conclusions to be drawn about the premium people place, for example, on the deaths of children compared to the over 60s, or on deaths where the victims are not to blame versus deaths where the victims are to blame.

To our knowledge, the research reported in this paper is the first to attempt such a quantification. A discrete choice experimental design was implemented which enabled us to estimate the “bad death premium” for four underlying features or attributes: the age of the victim, who was most to blame for the death, the severity of the victim’s pain and suffering in the period leading up to death, and the duration of the victim’s pain and suffering in the period leading up to death. These last two attributes provided a specific operationalisation of the “dread” factor based on Sunstein’s (1997) argument that especially dreaded deaths from diseases like cancer and AIDS deserve special attention in accordance with the degree of pain and suffering that precedes them. It should be noted however that these attributes may not fully capture other types of uneasiness or “dread” that might be associated with certain sudden, unanticipated, deaths like for example airplane crashes or terrorist attacks.

In addition, it may be that the ‘badness’ of a particular death depends not only on its generic characteristics (or attributes), but on the exact cause of death – or ‘context’. Any marked differences in the perceived badness of a death when the context was known, would suggest that we cannot estimate a bad death premium based on our knowledge of the underlying attributes (i.e. ‘generic’ information) alone. Although not the focus of the current paper, one aim of the study reported here was to explore the impact of contextual information on the perceived badness of deaths.

2. Methodology

2.1. The DCE Study

Discrete choice modelling has its origins in marketing and consumer research where it has been employed to establish the relative importance of different characteristics or

attributes in the demand for different goods or services (Louviere, Hensher & Swait, 2000). It is also being used increasingly in environmental valuation and health economics (Boxall, Adamowicz, Swait et al., 1996; Hanley, Wright & Adamowicz, 1998; Ryan & Hughes, 1997; Ryan & Gerard, 2003).

In a typical discrete choice experiment (DCE) study, individuals are presented with -and then asked to choose between -hypothetical goods or services involving different levels of attributes identified as being important. The method is rooted in both random utility theory (McFadden, 1973; Hanemann, 1984) and Lancaster's economic theory of value (Lancaster, 1966) which assumes that the choices people make are utility maximising and can be decomposed into the constituent utility bearing characteristics or attributes of the goods. As the assumptions inherent in the random utility model have been detailed extensively elsewhere, no further details are given here (see Louviere et al, 2000; Ryan & Hughes, 1997; Ryan & Gerard, 2003).

2.2. Attributes and Levels

As noted in the introductory section, four attributes were identified as potentially important factors in determining the "bad death premium". To recap, these attributes were: the age of the victim (*age*), who was most to blame for the death (*blame*), the severity of the victim's pain and suffering in the period leading up to death (*severity*), and the duration of the victim's pain and suffering in the period leading up to death (*duration*). Having selected the attributes, the second stage in the design was to choose the attribute levels. Levels were chosen that covered a plausible range of values for each attribute and were distinct enough from each other to make trade-offs possible (see Table 1).

The *age* attribute was defined in terms of the age of the typical victim. The four levels spanned the whole age range from the under 17s, 17-40s, 40-60s and over 60s. The *blame* attribute was defined in terms of who was most to blame for the death. The four levels were designed to cover situations where nobody in particular was to blame, to the individuals themselves, other individuals, and business or government. The *severity* of

pain and suffering attribute was defined in terms of the extent to which the victim's quality of life was affected in the period leading up to death. The two levels described the victim's quality of life as either a bit worse than normal or a lot worse than normal. The *duration* of pain and suffering attribute was defined in terms of the length of time that the victim's quality of life was affected in the period leading up to death. The four levels ranged from a few minutes, a couple of weeks, a year or two, and 3-5 years. One feature of the model we set out below distinguishes it from those typically used in DCE studies: a fifth attribute, numbers of deaths was included in the design to provide a *quantitative* scale against which to measure the "bad death premium" associated with changes between the levels of the other four attributes of interest (outlined in detail below)¹. Hence, a number of *deaths* variable was added with four levels; 10, 15, 25, and 50. These attributes and levels produce 512 different combinations.

TABLE 1 ABOUT HERE

2.3. The experimental design

Having identified the attributes and their levels the decision of what combinations to present as choices to respondents then arises. As a full factorial design would contain descriptions of all 512 possible combinations, this was not considered to be a practical option and some fractional factorial design was required. A fractional factorial design was constructed that involved a sub-set of 64 scenarios paired to give 32 choices. The design followed a 'foldover' methodology which allows for main effects to be estimated and which controls for 2-way linear interactions (see Louviere et al, 2000). The 32 pairs from the DCE design were then divided between three Versions of the questionnaire, with two pairs being common across all three. An example of one of the choices used in the design is given on the next page. An important feature of the 32 choices that made up the DCE design was that levels on all 5 attributes varied simultaneously each time.

¹ It should be noted, that other DCEs have used a 'cost' attribute as scale against which to measure the utility of scenarios, based on the WTP method (see, for example, Skjoldborg & Gyrd-Hansen, 2003. Ryan & Wordsworth, 2000).

Which is worse?

	A	B
Number of people who die	15 deaths	25 deaths
Age-group	Under 17 year olds	17-40 year olds
Quality of life in period leading up to death	A bit worse than normal for last 1-2 years of their lives	A lot worse than normal for last few weeks of their lives
Who is most to blame	Business or government	Nobody in particular

What do YOU think? A is much worse than B A is slightly worse than B B is slightly worse than A B is much worse than A
 (tick one)

2.4. Overview of questionnaire design

The questionnaire was divided into two main sections; generic and contextual. In total, the generic section consisted of 23 questions in the above format. The first five of these were ‘practice’ questions. For each practice question the attributes of both scenarios were the same except for one item. Each of the five practice questions then varied a different attribute. As well as familiarising respondents with the choice task, three of these questions provided a test of dominance. All else equal, we would expect that increasing the severity and duration of suffering and the number of deaths would add to the badness – or disutility -of the scenario².

Questions 6 to 23 were made up as follows. Twelve of them were part of the DCE design: two (Q10 & Q18) were common to all three Versions and the other ten were unique to a

² The two remaining attributes, blame and age do not yield obvious ‘dominance’ tests. Whilst we may hypothesize about the direction of preferences in each case, there is no ‘logical’ response to these questions.

particular Version. Splitting the questions between versions in this way gave coverage of all 32 pairs (i.e. 2 common questions and 30 version specific questions).

The other six questions in the generic section involved five pairs that would appear again in the contextual section (as Q24 to Q28); but here the contextual information was omitted and only the information about the underlying attributes given. One of these five pairs was presented twice – first as Q6, then again as Q21 – in order to check stability of responses. These six questions were not part of the DCE design and were included for two reasons. Firstly, these questions provided a between-sample test of sensitivity to the number of deaths; in moving from Versions 1 to 3 the number of deaths in scenario A was increased from 10 to 15 and then to 25, whilst the number of deaths in B remained at 10 in all three Versions. As the number of deaths in A becomes progressively higher we expect the proportion of respondents perceiving the deaths in B as worse than those in A to be smaller in Version 2 than in Version 1 and smaller still in Version 3. Secondly, these questions were used to see how well our estimated model could predict patterns of response to other generic scenarios that were not part of the estimation procedure.

The Repeated Question (Q6/Q21): number of deaths in Versions 1 (2, 3)

	A	B
Number of people who die	10 deaths (15, 25)	10 deaths (10, 10)
Age-group	17-40 year olds	17-40 year olds
Quality of life in period leading up to death	A bit worse than normal for last few minutes of their lives	A lot worse than normal for last few minutes of their lives
Who is most to blame	The individuals themselves	Business or government

In the contextual section – Q24 to Q28- respondents were presented with 5 choices each identical to one that appeared in the generic section, but with the context now revealed. For example, Q24 was identical to the repeated question (Q6/Q21) in the generic section except that the deaths were now described as being ‘of car drivers’ and ‘of rail passengers’ in scenarios A and B respectively. Hence, a comparison of the generic and contextual responses involves a within-subject comparison of responses. Table 2 shows the full range of questions that appeared in the contextual section. These questions were used to examine the extent to which inclusion of contextual information affected responses. As the impact of context on responses is reported in detail elsewhere (ref Anne’s paper), this aspect of the design is dealt with only briefly here.

TABLE 2 ABOUT HERE

2.5. Data Collection

A total of 313 respondents were recruited by a professional market and social research organisation (NWA) on a quota-sampling basis to be broadly representative of the socio-demographic characteristics of the UK population. One hundred and fifty four respondents (49.2%) were male and the average age was 44, (range 17 to 87). The data were collected in focus groups of between 8 and 12 respondents moderated by the authors. Groups were convened in Cambridge, Darlington, Durham, Norwich and Stockton in the UK.

At the beginning of each group the moderator provided the participants with a handout which explained the general purpose of the study and real-world examples used to illustrate the ways in which various types of deaths might differ on the attributes of interest.

Respondents were then presented with the generic section and taken question-by-question through the first five choices each of which varied the level of only one of the five attributes. The aim of these questions was to test whether respondents understood the

choice task and also provided the group with the opportunity to briefly discuss each attribute in turn. When they had completed the first five questions respondents were instructed to work through the remaining choices at their own pace. When all group members had completed the generic section of the questionnaire there was an opportunity for a brief discussion.

The generic section was then collected and the contextual section handed out. Respondents' attention was drawn to the fact that these questions now gave contextual information and that there was an invitation not only to tick a box but also to write a sentence or two in support of their decision. Finally, respondents were asked to complete a questionnaire requesting basic demographic information along with details of existing illness in the household

3. Results

Before presenting the results of the DCE model, we turn to the consistency checks that were built into the study design.

3.1. Consistency Checks

The first type of consistency checks were tests of dominance. These tests were provided by three of the first five choices that varied only one of the attributes. The percentages of respondents passing the dominance tests were 83.4% for the *duration* attribute, 92.9% for the *severity* attribute, and 92.3% for the *deaths* attribute. Hence, it did appear as if the vast majority of respondents understood the questions and were 'consistent' in their responses.

The second type of consistency check was test-retest reliability. This test was provided by presenting identical scenarios in the 6th and 21st choices. The aggregate data showed that 75.9% of respondents thought B was slightly worse or much worse than A when the scenario was presented as the 6th choice increasing slightly but not significantly to 80.1% when it was presented as the 21st choice ($\chi^2(1)=1.605, p=.20$). We used McNemar tests to

look at this data in more detail. Briefly, this is a non-parametric sign test for matched pairs of observations and tests whether there is any systematic pattern in those respondents changing responses between the 6th and 21st choices.

Table 3 shows the patterns of choices in Version 1 of the questionnaire (in which the numbers of deaths attribute had a level of 10 for both scenarios A and B). This shows that the overall split between A and B is fairly stable – 15:86 in Q6, 14:87 in Q21 – but those identifying B as worse become less extreme in their view: 61 regarded B as much worse than A in Q6, but this number falls to 49 in Q21. The McNemar test shows no significant difference at the 5% level - $\gamma^2(6)=5.13$, $p=.527$. Although the data is not shown here, the corresponding test statistics for Versions 2 and 3 are: $\gamma^2(6)=9.58$, $p=.143$ and $\gamma^2(6)=6.82$, $p=.337$ respectively. Hence, there are no systematic differences between responses to the two questions, indicating a good level of test-retest reliability. For the sake of parsimony, responses to Q21 only are presented below.

TABLE 3 ABOUT HERE

The third type of consistency check is the between-sample test of the sensitivity to the number of deaths. Table 4 shows the percentages making each response in the three Versions. Recall that in moving from Versions 1 to 3 the level of the number of deaths in scenario A increased from 10 to 15 and then to 25, while the number of deaths in B remained at 10 in all three Versions. As the number of deaths in A becomes progressively higher we expect the proportion of respondents perceiving the deaths in B to be worse than those in A to be smaller in Version 2 than in Version 1 and smaller still in Version 3. Table 4 shows that, as the number of deaths in scenario A was increased from 10 to 15 and then to 25 the percentages of respondents rating B as slightly worse OR much worse than A reduced from 86.1%³ in version one to 83.3% in version two to 70.4% in version three. A Pearson's chi squared test for independent samples shows that responses are significantly different across the three versions $\gamma^2(6)=18.57$, $p =.005$.

³ This figure is derived by summing the last 2 columns of table 3 i.e. 37.6% + 48.5%= 86.1% Ditto for versions 2 & 3.

TABLE 4 ABOUT HERE

The final type of consistency check built into the overall study design was a test of the impact of context on responses, results that are dealt with in more detail elsewhere (ref Anne's paper). Briefly, the results were mixed. For example, in the case of accidents at work vs car drivers we found no significant differences between generic and contextual responses when this pairing appeared as Q26 in Version 2 (10 vs 15 $W=-1.47$, $p=0.14$) and Version 3 (25 vs 50 $W=-0.37$, $p=0.72$). Significant differences were, however, found in a number of the other pairs considered. In the case of car drivers vs pedestrians scenarios we found significant differences between the generic and contextual responses both when that pairing appeared as Q26 in Version 1 (25 vs 15 $W=-4.61$, $p=0.000$) and as Q28 in Version 3 (15 vs 25 $W=-4.065$, $p=0.000$). Further details of this aspect of the study are available from the authors.

3.2. The DCE model

It is important to note that none of the results presented thus far involved those 32 pairings on which the DCE estimation procedure was based. We turn to those results now.

A logit model was used which estimates the probability that one scenario in a pair is considered to be worse than the other based on the levels of attributes in each. As logit is a binary choice model, we combined response modes 'slightly worse than' and 'much worse than' in the estimation process. On this basis, there was no significant difference in response to those questions – Q10 ($\chi^2(2)=5.82$, $p=.054$) and Q18 ($\chi^2(2)=2.417$, $p=.299$) – that were common to all three questionnaires. Hence, we considered it appropriate to pool the data across the three versions of the questionnaire.

If we assume that the disutility of the particular type of death, $U(X)$, is a function of the *age* of the victim, the *severity* and *duration* of any period of ill-health prior to death, and the primary *responsibility/ blame* for the death (i.e., $U(X) = f(\text{age}_X, \text{severity}_X, \text{duration}_X,$

blame_x), then the disutility of each combination of attributes and levels would be the disutility of the particular type of death multiplied by the numbers of deaths specified. Standard DCE estimation procedures, however, rely on taking *absolute differences* of attribute levels. We therefore estimated a model that was multiplicative in the number of deaths as follows:

$$P(B) = f\{(N^{\gamma}_B - N^{\gamma}_A) + \gamma_{age}(N^{\gamma}_B * age_B - N^{\gamma}_A * age_A) + \gamma_{severity}(N^{\gamma}_B * severity_B - N^{\gamma}_A * severity_A) + \gamma_{duration}(N^{\gamma}_B * duration_B - N^{\gamma}_A * duration_A) + \gamma_{blame}(N^{\gamma}_B * blame_B - N^{\gamma}_A * blame_A) + e\}$$

where $P(B)$ is the probability that a respondent will consider scenario B to be worse than A, N^{γ}_A and N^{γ}_B are the number of deaths in scenarios A and B respectively raised to power γ , and γ_i is the coefficient on the i^{th} attribute. When γ is set equal to one, all deaths are given equal weight (i.e., 50 deaths would be given five times the weight of 10 deaths). Values of γ less than 1 indicate a declining marginal disutility of deaths (i.e., 50 deaths would be given *less* than five times the weight of 10 deaths); while values of γ greater than 1 indicate an increasing marginal disutility of deaths. The remaining attributes enter the model as dummy variables with the omitted dummies representing the following base case: *age* (over 60s), *severity* (bit worse than normal), *duration* (last few minutes), *blame* (nobody in particular).

The disutility of this ‘base case’ of death was accorded a value of 1. We first estimated the model with $\gamma = 1$, but a number of the coefficients appeared to have the wrong sign and did not appear to fit the data at all well. It seemed that setting $\gamma = 1$ imposed a restriction on the weight placed on the number of deaths which diverged from what respondents actually did to such an extent that it distorted many of the other parameter estimates.

We then explored other values of γ . A grid-search showed that the log-likelihood function was minimized (i.e. the multiplicative model fitted best) when $\gamma = 0.2$. The results of estimating the model with $\gamma = 0.2$ are shown in Table 5. Observations were not

independent (as each respondent contributed 12 observations), so standard errors were adjusted to allow for clustering by respondent.

TABLE 5 ABOUT HERE

In our estimation the *deaths* variable is ‘offset’ to make the disutility of the base case type of death equal to 1 and so no figure is given for this variable in table 5. The remaining variables in table 5 are prefixed by a ‘D*’ to indicate that they are multiplicative in the number of deaths. With $\gamma = 0.2$ the coefficients shown in Table 5 are all signed in accordance with prior expectations. The dummies on *age* show that disutility increases as the age of the typical victim falls- i.e. deaths of younger people are worse, all other things being equal. Also in line with expectations, the dummy for *severity* ‘lot worse than normal’ (compared with ‘bit worse than normal’) increases the disutility of a scenario and disutility increases with duration of suffering. The dummies relating to *blame* show an interesting pattern. According to the model the dummy for blame ‘individuals themselves’ shows *reduced* disutility relative to the base case ‘nobody in particular’. On the other hand, the dummies for ‘other individuals’ and ‘business/ government’ both *increased* the disutility of a scenario.

We turn now to how well each model predicts responses to those generic questions that were not used in the estimation process. Again focusing on responses to Q21 (the generic equivalent of the car driver/rail passenger comparison), Table 4 showed that, as the number of deaths in scenario A was increased from 10 to 15 and then to 25 the percentages of respondents rating B as slightly worse OR much worse than A reduced from 86.1%⁴ in version one to 83.3% in version two to 70.4% in version three. The equivalent percentages predicted by the model are 85.6%, 83.3% and 79.5% in versions one, two and three respectively, not markedly different from the actual data. However, the model does appear to *underestimate* the sensitivity to the number of deaths. Consider

⁴ This figure is derived by summing the last 2 columns of table 3 i.e. 37.6% + 48.5%= 86.1% Ditto for versions 2 & 3.

that, in moving from Version 1 to Version 3, the model predicts the numbers rating B as worse than A to fall by 6.1% whereas the actual data showed a fall of 15.7%⁵.

This underestimation led us to fit a model with $\gamma = 0.3$ which was found to be a better fit of the data from the 21st choice (coefficients shown in Table 6). The predicted percentages of respondents rating B as slightly worse or much worse than A were 82.4% in Version 1, 78.0% in Version 2 and 70.5 % in Version 3 (i.e., a difference between versions one and 3 of 11.9%). In general, the model with $\gamma = 0.3$ performed rather better than that with $\gamma = 0.2$ at predicting responses to the remaining pairings not reported here. Even greater sensitivity of the fitted values would have been desirable in a number of cases; but, raising γ much further tended to destabilize other parameter estimates⁶. Whilst the model with $\gamma = 0.3$ appears to be the best fit of our ‘out of sample’ data, we focus below on the model with $\gamma = 0.2$ which was the best fit of the data used in the estimation procedure.

4. Discussion

The model presented above sets out to provide a ‘generic’ tool that may be used to estimate the relative badness of types of deaths that may be described in terms of levels on the four attributes. For example, we may use the coefficients on the model to estimate the disutility- or badness- of one death of a particular type by simply adding up the scores on the coefficients, allowing the disutility of the ‘base type’ of death⁷ to take on a value of 1.

For example, suppose we wished to compute the relative badness of two deaths which differed only in terms of who was most to blame for the deaths, holding other attributes

⁵ This is derived by subtracting 70.4% from 86.1%=15.7%.

⁶ Setting $\gamma = 0.4$ resulted in a number of coefficients taking on the ‘wrong’ sign and a dramatic fall in the value of the log-likelihood.

⁷ Recap, the ‘base type’ of death is: age (over 60s), severity (bit worse than normal), duration (last few minutes), blame (nobody in particular.)

constant. Consider two types of deaths- A and B- both affecting people over 60 whose quality of life would have been a lot worse than normal for the last 1-2 years of their lives, only differing in terms of who is most to blame for the deaths –business or government in the case of death A and the individuals themselves in the case of death B. By adding up the scores on the relevant coefficients from the $\gamma = 0.2$ model we would get:

$$U_{death\ A} = 1 + 0 + 0.331 + 0.276 + 0.537 = 2.144$$

$$U_{death\ B} = 1 + 0 + 0.331 + 0.276 + -0.259 = 1.348$$

where 1 is the disutility of the ‘single death’ base case, 0.331 is the coefficient on *severity* (lot worse than normal), 0.276 is the coefficient on *duration* (1-2 years), and 0.537 and -0.259 are the coefficients on *blame* (business/ government and individuals themselves respectively): the zeros included in the expressions reflect the fact that ‘age’ was the same as for the base case. The relative badness or disutility ratio between the two types of deaths can then be estimated:

$$\frac{U_{death\ A}}{U_{death\ B}} = \frac{2.144}{1.348} = 1.59$$

The corresponding estimate from the model in which $\gamma = 0.3$ is 1.73.

Consider another two deaths- C and D -both involving quality of life a bit worse than normal for the last few minutes of their lives, but which differ terms of the *age* of victims (the over 60s in C and 17-40s in D) and *blame* (other individuals in C and individual themselves in D). By the same method as above⁸, this example would then yield a ratio of 1.107 using the $\gamma = 0.2$ model.

$$U_{death\ C} = 1 + 0.331 + 0.526 = 1.857$$

$$U_{death\ D} = 1 + 0.605 + 0.331 + -0.259 = 1.677$$

⁸ In each case the ‘duration’ variables attract scores of zero as they are the same as the base case. By the same token, age also attracts a score of zero in the $U_{death\ C}$ calculation.

$$\frac{U_{deathC}}{U_{deathD}} = \frac{1.857}{1.677} = 1.107$$

The corresponding estimate from the model in which $\gamma = 0.3$ comes out at 1.85.

On the face of it, such a model would allow decision makers to simply ‘plug in’ the relevant characteristics of any two types of death in order to estimate the relative badness of each. Further, the impact of varying a level on any one attribute may be assessed directly (e.g. changing the ‘blame’ attribute from ‘individuals themselves’ to ‘other individuals’) within such a tool. Such a tool would be useful to policy makers attempting to quantify a bad death premium. There are, of course, a number of caveats to using the model in this way, which we set out below.

Firstly, it may be the case that deaths cannot simply be described in generic terms and that respondents may wish to consider additional ‘contextual’ information. That is, in some cases, the specific cause of death may matter *over and above* knowledge of its generic characteristics. Recall that significant differences between generic and contextual responses in a number of the pairs considered here. This could well reflect the fact that some influential variable(s) were omitted from the DCE model, or that being given information about context caused certain categories of included variables to be interpreted somewhat differently. Such issues require further examination before any generic model may be used to estimate the badness of particular causes of death. Rather, generic models ought to be interpreted as giving a broad indication of the types of factors that matter to respondents and the likely importance of those factors, all else equal.

Secondly, the estimated model showed marked insensitivity to the number of deaths and there are substantial – and, arguably, rather uncomfortable – implications of estimating γ at 0.2. In particular, rather than 50 deaths being five times as bad as 10 deaths, estimating $\gamma = 0.2$ implies that respondents regard 50 deaths as being just 1.37 times as bad as 10 deaths. Whilst a degree of diminishing marginal utility (or disutility in this case) might be expected under standard economic theory, it just seems implausible that

50 deaths are only 1.37 times as bad as 10 deaths. Even when γ is set = 0.3 (the model which performed better at predicting responses to those generic questions not used in the estimation procedure), 50 deaths are still only 1.62 times as bad as 10 deaths.

So, what is the possible explanation of this insensitivity which arises both in the estimated model and, to a lesser extent, in the responses to those generic questions not used in the estimation procedure? One possible explanation for the apparent insensitivity to numbers, is that respondents simply classify levels on attributes as, for example, 'high', 'medium', 'low', 'very low' and disregard the numerical information. That is, 50 might be regarded as 'high', and 15 as 'low', but the weight given to the difference between 'high' and 'low' would not necessarily reflect the magnitude of the differences in either relative or absolute terms. This raises the possibility that, had different levels on that attribute been selected- say, 20, 40, 60 & 100- then 100 may have been regarded as 'high' and 40 as 'low' and the information processed exactly as before.

Similarly, in discussing problems associated with the inclusion of probabilistic information in DCEs, Lloyd (2003) argues that 'fuzzy trace theory' predicts that respondents will result in risks being coded as 'high' or 'low'. This theory, put forward by Reyna and colleagues (Reyna & Brainerd, 1991) predicts that people only extract the 'gist' of any numerical information and then code this as a qualitative representation of probability with no representation of the *magnitude* of the differences between risk levels. As it is well documented that respondents use simplifying heuristics when presented with complex decision tasks (see, for example, Payne & Bettman,), it seems plausible that qualitative coding may also take place when respondents are presented with other numerical attributes- such as the number of deaths variable used here.

As reliance on simplifying heuristics is known to increase as the task becomes more complex, then any tendency to 'code' numerical information in this way is likely to be greater when all attributes are varied simultaneously than when certain attributes are held constant across choices. As outlined in the methods section, the 32 choices that made up the DCE design varied all 5 attributes simultaneously, whilst the remaining generic

questions (the counterparts of the contextual questions) generally held at least one of the attributes constant across choices. This may offer some explanation of why the model – based on the 32 questions in the DCE design- appears to predict less sensitivity than the actual responses to the generic questions used in the ‘out of sample’ test of predictive ability.

If at least a proportion of respondents in DCE studies are found to be processing numerical information in this way, then this may have profound implications for the reliability of the estimated values. In particular, the estimated values will be dependent on the levels of the attributes selected for use in the design, a phenomenon that does not appear to have been explored extensively in the literature. A number of studies that include a ‘cost’ attribute in order to estimate WTP have uncovered some evidence that estimates are dependent on the levels of that attribute. In one study that set out to estimate women’s WTP for cervical screening programmes, estimates were found to be sensitive to variations in the ‘price’ attribute across sub-samples (Ryan & Wordsworth, 2000). One large scale study conducted in Denmark set out to elicit preferences over different health care systems and found the use of a wider cost range and inclusion of higher payments to be associated with increased WTP values across sub-samples (Skjoldborg & Gyrd-Hansen, 2003). Concerns about the choice of level on the cost attribute in WTP studies in health have also been expressed by Ratcliffe (2000).

Further research is needed in order to determine whether this problem is pervasive in DCE studies that use numerical information of some form of another (for example, risk levels, WTP amounts or number of deaths) and it would appear to be relatively straightforward to test. This will certainly involve between sample tests using different levels- and possibly ranges- of the numerical variable and testing whether utility estimates are robust across treatments. In-depth qualitative analysis may help to identify whether or not a proportion of respondents are using some sort of coding heuristic and whether this tendency is greater, the more complex the decision. If this is found to be the case, then this would clearly have implications for the use of study designs such as the one used here in which all attributes varied simultaneously.

5. General Conclusions

As we have discussed there are a number of caveats associated with using the model we have estimated in the way we propose. There are clearly a number of theoretical and methodological issues that need to be addressed before such a model can be confidently applied. So in that sense we are still some way from a robust and reliable quantification how much of a “bad death premium” people attach to different types of underlying features or attributes. This paper is however the first to attempt such an exercise and despite its limitations provides the academic community with a good starting point to take things further. Moreover, we highlight some issues surrounding the use of numerical attributes in DCE studies and recommend areas where further research is needed.

ACKNOWLEDGEMENTS

We are indebted to Brett Day for his help with the econometric analysis presented and to Peter Moffatt for providing additional econometric advice. The idea of estimating a multiplicative model grew out of discussions with Bob Sugden and Ian Bateman and we acknowledge their considerable input into the ideas expressed here. Jordan Louviere, Deborah Street and Emma McKintosh offered general advice on the design of DCE experiments and we are grateful to them.

References

- Boxall, P. C., Adamowicz, W. L., Swait, J., Williams, M., & Louviere, J. (1996). A comparison of stated preference methods for environmental valuation. *Ecological Economics*, 18(3), 243-253.
- Bryan, S. Roberts, T. Heginbotham, C and McCallum, A. QALY-maximisation and public preferences: results from a general population survey. *Health Economics* (2002): 11 : 679-693.
- Carlsson, F., Johansson-Stenman, O., & Martinsson, P. (2004). Is transport safety more valuable in the air? *Journal of Risk and Uncertainty*, 28(2), 147-163.

- Chilton, S., Covey, J., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N., & Spencer, A. (2002). Public perceptions of risk and preference-based values of safety. *Journal of Risk and Uncertainty*, 25(3), 211-232.
- Cropper, M. L., Aydede, S. K., & Portney, P. R. (1994). Preferences for life saving programs - How the public discounts time and age. *Journal of Risk and Uncertainty*, 8(3), 243-265.
- Hammitt, J. K., & Liu, J. T. (2004). Effects of disease type and latency on the value of mortality risk. *Journal of Risk and Uncertainty*, 28(1), 73-95.
- Hanemann, W. M. (1984). Welfare evaluations in contingent valuation experiments with discrete responses. *American Journal of Agricultural Economics*, 66(3), 332-341.
- Hanley, N., Wright, R. E., & Adamowicz, V. (1998). Using choice experiments to value the environment - Design issues, current experience and future prospects. *Environmental & Resource Economics*, 11(3-4), 413-428.
- Jones-Lee, M.W., Hammerton, M., & Philips, P.R. (1985). The value of safety: Results of a national sample survey. *Economic Journal*, 95, 49-72.
- Lancaster, K. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74, 132-157.
- Lloyd, A.J. (2003) Threats to the estimation of benefit: are preference elicitation methods accurate?, *Health Economics*, 12, 393-402
- Louviere, J.J., Hensher, D.A., & Swait, J.D. (2000). *Stated Choice Methods: Analysis and Application*. Cambridge: Cambridge University Press.
- Mendeloff, J. M., & Kaplan, R. M. (1989). Are large differences in lifesaving costs justified - a psychometric study of the relative value placed on preventing deaths. *Risk Analysis*, 9(3), 349-363.
- McFadden, D. (1973). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed). *Frontiers in Econometrics*. New York: Academic Press (pp105-142).
- Ratcliffe J (2000), The use of conjoint analysis to elicit willingness to pay values. Proceed with caution? *Journal of Technological Assessment in Health Care*, 16 (1), 270-275.
- Ryan, M., & Gerard, K. (2003). Using discrete choice experiments to value health care programmes: Current practice and future research reflections. *Applied Health Economics and Health Policy*, 2(1), 1-10.
- Ryan, M., & Hughes, J. (1997). Using conjoint analysis to assess women's preferences for miscarriage management. *Health Economics*, 6(3), 261-273.
- Ryan M, & Wordsworth S (2000). Sensitivity of willingness to pay estimates to the level of attributes in discrete choice experiments. *Scottish Journal of Political Economy* 47: 504-524
- Reyna V, Brainerd C, (1991) Fuzzy-trace theory and framing effects in choice: gist extraction, truncation and conversion. *Journal of Behavioural Decision Making*, 4, 249-262.
- Savage, I. (1993). An empirical-investigation into the effect of psychological perceptions on the willingness-to-pay to reduce risk. *Journal of risk and uncertainty*, 6(1), 75-90.
- Skjoldborg, U.S, Gyrd-Hansen, D (2003) Conjoint analysis. The cost variable: an achilles heel? , *Health Economics*, 12: 479-491

- Subramanian, U., & Cropper, M. (2000). Public choices between life saving programs: The tradeoff between qualitative factors and lives saved. *Journal of Risk and Uncertainty*, 21(1), 117-149.
- Sunstein, C. R. (1997). Bad deaths. *Journal of Risk and Uncertainty*, 14(3), 259-282.

Table 1: Attributes and levels included in the DCE

<u>Attributes</u>	<u>Levels</u>
<i>Age range of typical victim</i>	<ul style="list-style-type: none"> • Under 17 years • 17-40 years • 40-60 years • Over 60
Who is most to <i>blame</i>	<ul style="list-style-type: none"> • Nobody in particular • The individuals themselves • Other individuals • Business/ government
<i>Severity of pain and suffering in period leading up to death</i>	<ul style="list-style-type: none"> • quality of life a bit worse than normal • quality of life a lot worse than normal
<i>Duration of pain and suffering in period leading up to death</i>	<ul style="list-style-type: none"> • a few minutes • a couple of weeks • a year or two • 3-5 years
Numbers of <i>deaths</i>	<ul style="list-style-type: none"> • 10 • 15 • 25 • 50

Table 2: The contextual scenarios

Question	Version 1	Version 2	Version 3
Q24	10 car drivers vs 10 rail passengers	15 car drivers vs 10 rail passengers	25 car drivers vs 10 rail passengers
Q25	50 smoking cancer vs 25 asbestos cancer	25 CO poisoning vs 15 accidents at work	15 smoking cancer vs 10 asbestos cancer
Q26	25 car drivers vs 15 pedestrians	10 accidents at work vs 15 car drivers	25 accidents at work vs 50 car drivers
Q27	15 pedestrians vs 25 breast cancer	25 pedestrians vs 15 breast cancer	15 CO poisoning vs 25 accidents at work
Q28	10 work-related cancer vs 15 car drivers	25 work-related cancer vs 50 car drivers	15 car drivers vs 25 pedestrians

Table 3: Frequencies of responses for the repeated question (Q6/Q21) in Version 1

Responses to Q6					
Responses to Q21	A much worse than B	A slightly worse than B	B slightly worse than A	B much worse than A	Total
A much worse than B	1	2	1	1	5
A slightly worse Than B	2	4	2	1	9
B slightly worse Than A	2	2	13	21	38
B much worse than A	1	1	9	38	49
Total	6	9	25	61	101

Table 4: Responses to Q21 (as percentages) in Versions 1, 2 and 3

	A much worse than B	A slightly worse than B	B slightly worse than A	B much worse than A
Version 1 (deaths _A =10, deaths _B =10)	5.0%	8.9%	37.6%	48.5%
Version 2 (deaths _A =15, deaths _B =10)	3.7%	13.0%	53.7%	29.6%
Version 3 (deaths _A =25, deaths _B =10)	11.2%	18.4%	40.8%	29.6%

Table 5: The DCE model with $\alpha = 0.2$

Dummy Variable	Coefficient	Robust Std Error	Significance
Deaths	offset		
D*age(<17s)	.817	.065	<.001
D*age(17-40s)	.605	.055	<.001
D*age(40-60s)	.355	.049	<.001
D*severity(lot worse than normal)	.331	.035	<.001
D*duration(couple of weeks)	.152	.045	.001
D*duration(1-2yrs)	.276	.047	<.001
D*duration(3-5yrs)	.354	.053	<.001
D*blame(individuals themselves)	-.259	.057	<.001
D*blame(other individuals)	.526	.054	<.001
D*blame(business/government)	.537	.051	<.001
Log pseudo-likelihood	-1839		
Number of observations	3401		

Table 6: The DCE model with $\alpha = 0.3$

Dummy Variable	Coefficient	Robust Std Error	Significance
Deaths	offset		
D*age(<17s)	.526	.051	<.001
D*age(17-40s)	.293	.041	<.001
D*age(40-60s)	.121	.032	<.001
D*severity(lot worse than normal)	.196	.027	<.001
D*duration(couple of weeks)	.096	.034	.004
D*duration(1-2yrs)	.175	.040	<.001
D*duration(3-5yrs)	.316	.051	<.001
D*blame(individuals themselves)	-.231	.047	<.001
D*blame(other individuals)	.295	.043	<.001
D*blame(business/government)	.345	.036	<.001
Log pseudo-likelihood	-1886		
Number of observations	3401		

**This working paper has been produced by
the Department of Economics at
Queen Mary, University of London**

**Copyright © 2007 Angela Robinson, Judith Covey,
Anne Spencer and Graham Loomes. All rights reserved**

**Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/papers/wp.htm**