

## Durham Research Online

---

### Deposited in DRO:

15 July 2010

### Version of attached file:

Accepted Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Robinson, A. and Covey, J. and Spencer, A. and Loomes, G. (2010) 'Are some deaths worse than others ? the effect of labelling on peoples perceptions.', *Journal of economic psychology.*, 31 (3). pp. 444-455.

### Further information on publisher's website:

<http://dx.doi.org/10.1016/j.joep.2010.01.011>

### Publisher's copyright statement:

NOTICE: this is the author's version of a work that was accepted for publication in *Journal of economic psychology*.

### Additional information:

## Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

## 1. BACKGROUND

People regard some deaths as worse than others (Sunstein, 1997). For example, research has shown people are more concerned about cancer deaths than deaths from heart disease, motor vehicle accidents, household fires, or airplane accidents (Jones-Lee et al, 1985; Savage, 1993). Likewise deaths caused by industrial air pollution are regarded as more deserving of resources than deaths caused by smoking or automobile accidents (Subramanian, 2000). Hence, it would appear that just as the acceptability of risks can be characterized by ‘qualitative’ factors, (Slovic, 1992) public concern about deaths might be aggravated by certain underlying features of those deaths.

As noted by Sunstein (1997) the risk perception literature does suggest reasons why people might regard some deaths as worse than others. More specifically, as well as the notion of livable life-years (i.e., it is worse if a child is killed than an older adult), he highlights the importance of ‘dread’ (i.e., death preceded by unusual pain and suffering), blameworthiness (i.e., responsibility for death lies with a third party), distributional equity (i.e., victims are members of socially disadvantaged groups), and high externalities (i.e., catastrophic events involving widespread non-pecuniary losses). For example, cancer deaths might be more ‘dreaded’ than deaths from heart disease because they are preceded by a longer period of pain and suffering, whereas people who die from industrial air pollution are less to blame than smokers are for their deaths.

The UK Department for Transport (DfT), currently use a willingness to pay (WTP)-based value for the value of preventing a fatality (VPF) on the roads of approximately £1.6 millions, but how transferable is this to other contexts which have different underlying attributes that are weighted more or less heavily? The Treasury Green book acknowledges that “there is

evidence that individuals are not indifferent to the cause and circumstances of injury or fatality”. It goes on to highlight that the Health and Safety Executive (HSE), in assessing the benefits of avoiding asbestos-related deaths- currently doubles the roads VPF figure to allow for individual aversion to dying from cancer as well as the additional associated personal and medical costs. They go on to say, however, that there is no direct evidence to support the magnitude of the adjustment factor used.

One way in which such policy decisions may be better informed is to quantify the importance of each of the underlying attributes of types of deaths, such as ‘dread’ and ‘blameworthiness’. This type of quantification would then allow more general conclusions to be drawn about the ‘bad deaths’ premium that people place on different types of deaths according to their underlying features. It is unclear, however, whether people’s concerns for different types of deaths *can* be adequately captured by a manageable set of ‘generic’ attributes used to describe the deaths.

There is much evidence from elsewhere to suggest ‘labelling’ the cause of death might affect responses. In the health state valuation field, Sackett and Torrance (1978) found that adding a label to a health state description significantly affected the utility values obtained:

‘tuberculosis’ was given a higher value than ‘unnamed contagious disease’, whereas ‘mastectomy for breast cancer’ was given a lower value than ‘mastectomy for injury’. Gerard et al (1993) also found differences when the word ‘cancer’ was used and when descriptions were written in the third party. Likewise, Rabin et al (1993) and Robinson and Bryan (2001) report that adding a label significantly affected valuations of both physical and mental conditions, but found differences in the direction of these effects. Smith (2008) found that

respondents were willing to pay more to avoid health states labeled as ‘stroke’ and ‘bowel cancer’ than their identical ‘generic’ counterparts.

The impact of assigning psychiatric ‘labels’ to individuals with mental health problems has been explored previously (Link et al., 1987; Loman & Larkin, 1976) and found to be important. For example, the use of a label such as ‘schizophrenia’ has been found to have strong effects on people’s perceptions and judgments about individuals with this mental illness (Link et al., 1987). Fryer and Cohen (1988) found that hospital staff rated patients labelled “psychiatric” as less likeable and as having more unfavourable traits than patients labelled “medical”.

In the current context, it is plausible that respondents’ attention to attributes of deaths might be directed by their affective reactions to the descriptions used. Just as research on the identifiable victim effect has shown (Jenni & Loewenstein, 1997; Kogut & Ritov, 2005; Small & Loewenstein, 2003), identifying the deaths by names may produce more of an emotional reaction than the stripped down generic descriptions. For, example Kogut and Ritov (2005) have found that people’s ratings of distress and willingness to contribute towards saving a single child’s life was higher when identifying information about the child was provided (i.e., their age, name and picture). By identifying the victim in this way the case becomes more vivid and concrete and evokes a stronger emotional response. However, it is worth noting that this identification effect may be restricted to single victims. When asked to rate their distress and willingness to contribute towards saving a group of eight children the same detail of identifying information about each of the eight children did not affect respondents’ willingness to contribute (Kogut and Ritov, 2005).

Whilst there are clearly a large number of potential factors that may influence respondents' assessments, the research published in this paper explored this issue for deaths that were described using four of Sunstein's attributes – the age of the victim (livable life-years), the severity and duration of the victim's pain and suffering in the period leading up to their death (dread), and who is most to blame for the victim's death (blameworthiness). More specifically, the aim of the study was to test how well people's concerns towards these attribute-only or 'generic' descriptions of deaths matched with their concerns towards descriptions where the specific cause of death is also identified (for example, as a driver in a car accident) – 'contextual' descriptions.

Whilst evidence that responses to generic and contextual descriptions are different would tend to rule out the existence of a generic model of 'bad deaths' which would be extremely useful for policy purposes, it does not, in itself, tell us that one description is somehow 'superior' to the other. Before reaching such a conclusion, we need to establish which set of responses is the most valid. One criterion against which to assess the validity of responses is to test the sensitivity to a factor that should have an impact on people's concern (Loomes, 2006). We might expect that, all other things being equal, people should regard more deaths as worse than less.

A prominent issue in the willingness to pay (WTP) literature, however, has been the inability of that method to account for insensitivity to the magnitude of the risk reduction. That is, respondents tend to view safety improvements as a 'good thing' and may therefore be liable to state much the same WTP for different sizes of risk reduction, whether for fatal or non-fatal injuries (Beattie et al., 1998; Covey et al., 1998; Dubourg et al., 1997; Hammitt and Graham, 1999; Jones-Lee et al., 1985; Jones-Lee and Loomes, 1995). It is worth noting that a number

of studies have uncovered marked insensitivity even when the risk reduction is couched in terms of the numbers dying (see for example, Beattie et al., 1998 and Desaignes and Rabl, 1995). By way of illustration, roughly half the sample in the Beattie et al. study (1998) stated exactly the same WTP amount for an improvement that prevented 15 deaths on the roads each year as one that prevented 5 deaths, even after this apparent anomaly had been pointed out to them and they had been given the opportunity to revise their responses.

As sensitivity to the magnitude of the risk reduction has become the ‘acid test’ of the validity and reliability of the WTP method (Carson, 1997; Yeung et al., 2003), it seems reasonable that other value elicitation techniques that set out to address similar issues are assessed against that same criterion. Hence, if we find that sensitivity to the numbers dying is significantly different when contextual descriptions are used rather than generic descriptions we would have an indication that the labelling has affected how the number attribute has been weighted in respondents’ choices.

In summary the aim of the research we focus on here was to explore the extent to which perceptions regarding the ‘badness’ of different types of deaths differ according to whether ‘generic’ or ‘contextual’ descriptions were used. Further, we set out to test whether sensitivity to the numbers of deaths differed across the ‘generic’ and ‘contextual’ versions of the questions.

## **2 METHODS**

### **2.1 Materials**

We elected to use a discrete choice experiment (DCE) in order to estimate the *generic* ‘bad deaths’ model based on characteristics of hazards such as the age of a typical victim, length of

illness or suffering preceding death, who is to blame for the death etc. In a typical discrete choice experiment (DCE) study, individuals are asked to choose between -hypothetical goods or services involving different levels of attributes identified as being important. Optimal experimental design methods exist that allow valid models to be estimated from a small subset of all possible combinations of attributes and levels. The principles of optimal study design, however, rely on all attribute levels varying independently, which limits the extent to which researchers are able to ‘set’ attribute levels according to scenarios of particular interest.

As the primary aim of the research reported here is to explore preferences over ‘generic’ and ‘contextual’ descriptions of hazards, *holding other factors constant*, questions from a DCE design had to be supplemented with others designed specifically for our purposes. These supplementary questions allowed us to ‘set’ attributes and levels according to scenarios of interest and explore the generic versus contextual issue in a controlled and systematic manner. Whilst it is the responses to these supplementary questions that are of primary interest here, it is necessary to outline briefly the DCE design in order to give a coherent account of the overall study design.

The questionnaire was in two parts: Part one presented the ‘generic’ (i.e. unlabelled) questions whilst the ‘contextual (i.e. labelled) questions appeared in Part two. A series of questions presented respondents with two premature death scenarios, labelled A and B, which varied on one or more of the five attributes shown below:

- The **numbers** of people who die (either 10, 15, 25, or 50)
- Their typical **age** (under 17s, 17-40s, 40-60s, or over 60s)
- How **much** their quality of life is affected in the period leading up to their deaths (either a bit worse than normal or a lot worse than normal)

- How **long** their quality of life is affected in the period leading up to their deaths (either a few minutes, a couple of weeks, 1-2 years, or 3-5 years)
- Who is most to **blame** for the deaths (the individuals themselves, other individuals, business/government or nobody in particular)

These attributes and levels produce a total of 512 different combinations. For the purposes of estimating a generic model using the DCE methodology, a fractional factorial design was constructed that involved a sub-set of 64 scenarios paired to give 32 choices (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. As above, the questions from the DCE design were supplemented by others to better address the specific issues of interest here.

**Figure 1a. Example of ‘generic’ question**

**Which is worse?**

	<b>A</b>	<b>B</b>
<b>Number of people who die</b>	50 deaths	25 deaths
<b>Age-group</b>	Over 60 year olds	Over 60 year olds
<b>Quality of life in period leading up to death</b>	A bit worse than normal for last 1-2 years of their lives	A lot worse than normal for last 1-2 years of their lives
<b>Who is most to blame</b>	The individuals themselves	Business or Government

What do YOU think?

(tick one)

A is much worse than B

A is slightly worse than B

B is slightly worse than A

B is much worse than A



**Figure 1b. Example of ‘contextual’ question**

**Which is worse?**

	<b>A</b>	<b>B</b>
<b>Number of people who die</b>	50 deaths from lung cancer caused by smoking	25 deaths from asbestos-related cancers
<b>Age-group</b>	Over 60 year olds	Over 60 year olds
<b>Quality of life in period leading up to death</b>	A bit worse than normal for last 1-2 years of their lives	A lot worse than normal for last 1-2 years of their lives
<b>Who is most to blame</b>	The individuals themselves	Business or Government

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)                                                                                       

In total, the generic part of the questionnaire consisted of 23 questions in the format depicted in figure 1a. In each question participants were asked to rate which of the scenarios they thought was the worst by ticking one of four possible responses, namely: A is much worse than B; A is slightly worse than B; B is slightly worse than A; or B is much worse than A. The first five questions were ‘practice’ questions in which the attributes of both scenarios were held the same except for one item. Each of the five practice questions then varied a different attribute.

Questions 6 to 23 were made up as follows. Twelve of them were part of a discrete choice experimental design. The other six generic questions involved five pairs that would appear again in Part 2 with contextual information; but in Part1 this contextual information was omitted and only the generic information was given. Two of these ‘generic’ questions were identical to one another and were presented as the 6<sup>th</sup> and 21<sup>st</sup> questions, providing a test-

retest reliability check on responses to ‘generic’ questions. This test was included to check that any observed differences were due to a genuine difference between the ‘generic’ and ‘contextual’ responses, and not an artifact of the questions being asked twice.

In Part 2, respondents were presented with the same 5 pairs they were presented with in Part 1, but this time the contextual information was included as shown in Figure 1b. Between the three versions of the questionnaire nine different causes of deaths were used to reflect the range of variation on the attributes (i.e., car drivers, pedestrians, rail passengers, cancer caused by smoking, cancer caused by asbestos, cancer caused in the workplace, accidents at work, breast cancer and carbon monoxide (CO) poisoning). As shown in Figure 2 these causes of deaths were paired up in a total of seven different ways with each ‘contextual’ pairings having an equivalent ‘generic’ pairing, identical other than the deaths being ‘unlabelled’ in the latter.

**Figure 2: Causes of death pairings used in each version of the questionnaire**

Cause of death pairing <sup>a</sup>	Version 1	Version 2	Version 3
A: Car Drivers	10 deaths	15 deaths	25 deaths
B: Rail Passengers	10 deaths	10 deaths	10 deaths
A: Pedestrians	15 deaths	25 deaths	
B: Breast Cancer	25 deaths	15 deaths	
A: Work-related Cancer	10 deaths	25 deaths	
B: Car Drivers	15 deaths	50 deaths	
A: Car Drivers	25 deaths		15 deaths
B: Pedestrians	15 deaths		25 deaths
A: Smoking Cancer	50 deaths <sup>b</sup>		15 deaths
B: Asbestos Cancer	25 deaths <sup>b</sup>		10 deaths
A: Accidents at Work		10 deaths	25 deaths
B: Car Drivers		15 deaths	50 deaths
A: CO Poisoning		25 deaths	15 deaths
B: Accidents at Work		15 deaths	25 deaths

<sup>a</sup>In each case, the ‘generic’ counterpart was identical to the ‘contextual’ question other than it being ‘unlabelled’

<sup>b</sup>Examples shown in Figure 1a (‘generic’ question) and 1b (‘contextual’ question)

As shown in Figure 2 each pairing was included in two of the three versions apart from car drivers vs. rail passengers which was included in all three versions. The number of deaths varied across versions and in some cases evidence from piloting suggested which way a weighting would go. For example, we knew from piloting that deaths from asbestos-related cancer would be given more weight than deaths from lung cancer attributed to the individuals themselves being smokers. So, in both versions the question showed that more deaths were caused by smoking-related cancer than asbestos-related cancer. However, as shown in Figure 2 the ratio between numbers of deaths was higher in version 1 (50:25 or 2:1) than it was in version 3 (15:10 or 1.5:1); likewise for work-related cancers vs. car drivers and for accidents at work vs car drivers. As our sensitivity to number of deaths tests are based on responses to such pairings, the remainder are not discussed here.

## **2.2. Sample/ Data Collection**

Data were collected by means of small discussion groups comprising between 8 and 12 participants. By using the services of a professional social and market-research company we recruited a quota sample of 313 people that was broadly representative of the gender, age and social class profile of the general population. Of the 313, 154(49.2%) were male, 159(50.8%) were female. The number of respondents aged 17 to 34, 35 to 54, and 55 to 90 was 121(38.7%), 97 (31%) and 94(30%) respectively. Two hundred and fifty three (80.8%) respondents had no long term health problem that limited their daily activity.

The group discussions began with a brief introduction to the aims of the study and participants were told that the Health and Safety Executive (HSE) wanted some guidance from members of the public about whether more effort and resources to be put towards

preventing some sorts of deaths than others. Part1- the 'generic' questionnaire -was then handed out and respondents taken through the 5 'practice' questions (in which one attribute at a time was varied), followed by a brief discussion to check for understanding. Respondents then worked through the remainder of the generic questions at their own pace with no further group discussion. Part 2- the contextual questionnaire- was then handed out and respondents worked through the questions at their own pace. A box was provided at the bottom of the page and respondents invited to write a sentence or two giving their reasons for their answers.

### **2.3 Analysis**

A brief account of the generic model estimated from responses to those questions that made up the DCE design is given in the Appendix. The analysis and reported in detail here relates only to those questions that-supplementary to the DCE design- that deal with the 'labelling' issue which is the focus of this paper.

In the 'generic' and 'contextual' questions the four response categories were scored 1-4, i.e., 1= A is much worse than B, 2= A is slightly worse than B, 3=B is slightly worse than A, 4=B is much worse than A. Both the test-retest reliability check and comparisons of the responses to the 'generic' and 'contextual' questions required within-subject tests of differences between responses. We therefore used Wilcoxon Signed Ranks tests to test whether the responses to the 'generic' and 'contextual' questions were significantly different from one another. To test whether responses were significantly different when different ratios of numbers of deaths were presented required between-subject tests. We therefore used either Kruskal-Wallis (when three versions were compared) or Mann-Whitney tests (when two versions were compared) to test whether the responses to each question were significantly different between the versions of the questionnaire.

Thematic qualitative analysis was conducted on the written reasons that participants gave when they answered the ‘contextual’ questions. Thematic analysis is one of the most commonly used methods of qualitative analysis. As outlined by Braun and Clarke (2006) the task of the researcher is to identify a limited number of themes which adequately reflect their data. Unlike content analysis in which a checklist of codes are applied to the data (Krippendorff, 1980), thematic analysts create their codes by defining what they see in the data. Codes emerge as the data is scrutinised. Hence, coding is a fluid process in which codes may be modified or altered as ideas develop. Themes which integrate sets of codes are then defined by the researcher and illustrated in the report with examples and, where necessary, numerical indications of the prevalence of each theme.

In the present paper this coding process was conducted by one of the current authors (JC). A detailed record was kept of the coding system and how it was linked to the original data using QSR Nud\*ist 4 software. The fit of this analysis to the original data was reviewed by other members of the research team (AR and AS)."

### **3. RESULTS**

#### **3.1 Test-retest Reliability Check**

Recall there was a repeated ‘generic’ question that was asked twice in Part 1 of the questionnaire at Q6 and Q21. Although the results showed that the variance in responses and proportions rating B as ‘much worse’ than A appear to be slightly lower for Q21 than Q6 (i.e., standard deviations 0.86 vs. 0.99; 35.6% vs. 41.2%), the Wilcoxon Signed Ranks test showed that there was no significant difference in the distributions of responses between Q6 and Q21. The test-retest reliability of the ‘generic’ questions was therefore of an acceptable level.

#### **3.2 Differences between responses to ‘generic’ and ‘contextual’ questions**

Table 1 summarises the results, aggregated across versions, for the ‘generic’ and ‘contextual’ questions for those pairings where the ratios of numbers were increased in the same direction in all versions of the questionnaire.

For each of the seven pairings from which we obtained responses from both ‘generic’ and ‘contextual’ questions we report the percentages of respondents giving ratings from 1 (A is much worse than B) to 4 (B is much worse than A), the means and standard deviations of the ratings, and the Z statistic obtained from the Wilcoxon Signed Ranks tests.

**Table 1: Comparison of responses to ‘generic’ and ‘contextual’ questions**

A vs. B	% Response				Mean	SD	Z <sup>a</sup> (p)
	1	2	3	4			
Car Drivers vs. Rail Passengers (N=306)							
Generic (Q21)	6.5	13.2	44.4	35.6	3.09	0.86	Z=3.48
Contextual	7.5	10.1	28.8	53.6	3.28	0.93	(p<.001)
Work-related Cancer vs. Car Drivers (N=205)							
Generic	31.7	36.6	20.5	11.2	2.11	0.98	Z=7.81
Contextual	61.5	28.8	8.8	1.0	2.49	0.70	(p<.001)
Smoking Cancer vs. Asbestos Cancer (N=201)							
Generic	20.4	20.4	30.8	28.4	2.67	1.10	Z=4.31
Contextual	14.4	12.4	21.4	51.7	3.10	1.10	(p<.001)
Accidents at Work vs. Car Drivers (N=206)							
Generic	37.4	39.8	16.5	6.3	1.92	0.89	Z=0.81
Contextual	42.2	38.3	10.7	8.7	1.86	0.93	(p=0.42)

Notes:

<sup>a</sup>Wilcoxon Signed Ranks Test

No significant differences in the distributions of responses were found for Accidents at work vs. Car Drivers – the majority regarding deaths from Accidents at Work as worse than Car Driver deaths was similar in the ‘generic’ and ‘contextual’ versions (77.2% vs. 80.5%). Significant differences were, however, found for the remaining 3 pairings for which these comparisons could be made.

Car Drivers vs. Rail Passengers: Although similar percentages of respondents regarded rail passenger deaths as worse than car driver deaths in both the ‘generic’ and ‘contextual’ versions (80.0% and 82.4%) the ‘contextual’ version produced a lower proportion of ‘slightly worse than’ (28.8% vs. 44.4%) and higher proportion of ‘much worse than’ responses (53.6% vs. 35.6%). Work-related Cancer vs. Car Drivers: Just over 20% more respondents regarded work-related cancer deaths as worse than car driver deaths in the ‘contextual’ version than the ‘generic’ version (90.3% vs. 68.3%). Smoking Cancer vs. Asbestos Cancer: About 10% more respondents regarded asbestos cancer deaths as worse than smoking cancer deaths in the ‘contextual’ version than the ‘generic’ version (73.4% vs. 69.2%) and the proportion of ‘much worse than’ responses were also higher (51.7% vs. 28.4%).

These findings lead us to the conclusion that people’s responses to ‘generic’ questions in which the causes of death are represented by attributes only are not completely predictive of their responses to ‘contextual’ questions which provide additional information about the specific causes of death.

### **3.3 Sensitivity to different ratios of deaths in the ‘generic’ and ‘contextual’ questions**

For these tests the ratios of deaths in the questions were increased between the versions of the questionnaire depending on which pairing of scenarios was used (see Figure 2).

- (i) 1:1 vs. 1.5:1 vs. 2.5:1 – Car Drivers vs. Rail Accidents

- (ii) 1.5:1 vs. 2:1 – Smoking Cancer vs. Asbestos Cancer; or 1:1.5 vs. 1:2 – Accidents at Work vs. Car Drivers and Work-related Cancer vs. Car Drivers,

Table 2 summarises the results for these comparisons. We report the percentages of respondents giving ratings from 1 (A is much worse than B) to 4 (B is much worse than A), the means and standard deviations of the ratings, and the chi-square ( $\chi^2$ ) or Z statistics obtained from the Kruskal-Wallis or Mann-Whitney U tests.

**Table 2: Sensitivity to ratios of deaths in ‘generic’ and ‘contextual’ questions**

A vs. B		% Response				Mean	SD	$\chi^2$ <sup>a</sup> or Z <sup>b</sup> (p)
		1	2	3	4			
Car Drivers vs. Rail Passengers								
Generic (Q21)	1:1	5.0	8.9	37.6	48.5	3.30	0.83	$\chi^2=11.69$ (p=.003)
	1.5:1	3.7	13.1	54.2	29.0	3.08	0.75	
	2:1	11.2	18.4	40.8	29.6	2.89	0.96	
‘contextual’	1:1	7.9	5.0	21.8	65.3	3.45	0.91	$\chi^2=8.19$ (p=.017)
	1.5:1	4.7	9.3	37.4	48.6	3.30	0.83	
	2:1	10.2	16.3	26.5	46.9	3.10	1.02	
Work-related Cancer vs. Car Drivers								
Generic	1:1.5	38.4	34.3	17.2	10.1	1.99	0.98	Z=1.87 (p=.062)
	1:2	25.5	38.7	23.6	12.3	2.23	0.97	
‘contextual’	1:1.5	69.7	23.2	6.1	1.0	1.38	0.65	Z=2.34 (p=.019)
	1:2	53.8	34.0	11.3	0.9	1.59	0.73	
Smoking Cancer vs. Asbestos Cancer								
Generic	1.5:1	14.1	15.2	33.3	37.4	2.94	1.05	Z=3.45 (p<.001)
	2:1	26.5	25.5	28.4	19.6	2.41	1.08	
‘contextual’	1.5:1	14.1	14.1	19.2	52.5	3.10	1.11	Z=0.05 (p=.96)
	2:1	14.7	10.8	23.5	51.0	3.11	1.10	
Accidents at Work vs. Car Drivers								
Generic	1.5:1	34.0	42.5	20.8	2.8	1.92	0.81	Z=0.62 (p=.54)
	2:1	41.0	37.0	12.0	10.0	1.91	0.96	
‘contextual’	1.5:1	44.3	38.7	11.3	5.7	1.78	0.86	Z=0.96 (p=.34)
	2:1	40.0	38.0	10.0	12.0	1.94	0.99	

Notes:

<sup>a</sup>Kruskall-Wallis Test, <sup>b</sup>Mann-Whitney U Test

The results show that although participants’ responses were equally sensitive or insensitive to the different ratios in both the ‘generic’ and ‘contextual’ questions in two of the four pairings



(Car Drivers vs. Rail Passengers, Accidents at Work vs. Car Drivers), sensitivity to the ratios was affected by providing information about the specific causes of death in the other two pairings. However, the way in which providing this information affected sensitivity was not the same. For one pairings the sensitivity seems to have been greater in the ‘contextual’ questions than the ‘generic’ questions (Work-related Cancer vs. Car Drivers), with the opposite result for the other pairing (Smoking Cancer vs. Asbestos Cancer).

Overall, there was therefore no systematic tendency for the responses to the ‘generic’ questions to be any more or less valid than those to the ‘contextual’ questions. So we cannot draw any definitive conclusion about which set of responses is by default the most appropriate.

However, these results also suggest that any conclusions we might draw about the relative importance that people place on the different attributes from the responses they gave to the ‘generic’ questions might be quite different from the conclusions we might draw from their responses to the ‘contextual’ questions. So if we take these results along with those reported in the previous section it appears that people’s responses to the ‘generic’ questions do not serve as reliable proxies for their responses to ‘contextual’ questions. The reasons for this mismatch were explored by analyzing the written reasons that participants gave when they answered the ‘contextual’ questions.

### **3.4 Results of the thematic qualitative data analysis**

Thematic analysis was conducted on the written reasons that participants gave when they answered the ‘contextual’ questions. We were particularly interested in gaining insight into the reasons why their responses to these questions were different to the ‘generic’ equivalents.

Hence, the analysis was only conducted on the three pairings shown in Table 2 which produced significant differences between the ‘generic’ and ‘contextual’ questions – focusing our attention on the reasons given by those participants who gave very different answers to both questions (i.e., rated A as worse than B in the ‘generic’ version and rated B as worse than A in the ‘contextual’ version – or vice versa).

### 3.4.1 Car Drivers vs. Rail Passengers

28 of the 34 respondents who switched to rating B (rail passenger deaths) as worse than A (car driver deaths) in the ‘contextual’ version gave written comments. The majority of the reasons given were linked to the ‘blame’ and ‘quality of life’ attributes that were provided in the ‘generic’ questions – i.e., that the individuals were not to blame or were helpless (14 cases – *“Not the individual’s fault”* [1005\_v1]; *“The individual is powerless to influence events”* [1048\_v3]), the greater suffering of the rail accident victims (6 cases – *“Due to the suffering that they had more than to who was to blame”* [2063\_v3]), or the violent nature of the death in a rail accident (2 cases – *“In rail accidents parts of bodies are found everywhere along the railway lines”* [1023\_v2]). However, a notable minority of reasons provided which seemed to be beyond the scope of the attributes presented the ‘generic’ questions – issues that expressed the view that deaths should just not be happening on the railways – i.e., the railways should be safe (10 cases *“Whilst all travel involves risk, when traveling by train danger is not expected”* [1057\_v3]), and rail accidents should be avoidable (2 cases – *“Avoidable by better maintenance of track and carriage”* [2044\_v3]).

23 of the 27 respondents who switched to rating A (car driver deaths) as worse than B (rail passenger deaths) in the ‘contextual’ version gave written comments. The larger numbers of deaths on the roads was mentioned by 9 respondents (*“The greater amount of car deaths is*

*the factor here*” [1036\_v3] ). Rail accidents were also perceived as relatively rare (4 cases – *“Rail accidents are much rarer”* [1073\_v1]), and trains were regarded as a safe method of travel (3 cases – *“Train travel is much safer than road”* [2013\_v2]). Other reasons highlighted how the context can raise issues that may be specific to the individual. For example, the importance of sorting out certain types of bad driver behavior (3 cases – *“I think more resources should be put into stopping speeding although I think B is actually worse for the people concerned* [2094\_v3]), or personal experience (2 cases – *“Having been involved in a car accident recently I am leaning towards this”* [1067\_v2]).

#### 3.4.2 Smoking Cancer vs. Asbestos Cancer

46 of the 51 respondents who switched to rating B (asbestos cancer) as worse than A (smoking cancer) in the ‘contextual’ version gave written comments. All of the themes that emerged were linked to the ‘blame’ attribute – i.e., that smokers only have themselves to blame whereas the victims of asbestos related cancer were not at fault (29 cases *“Smoking is a choice so if they are dying from smoking related cancer then its only themselves to blame”* [1003\_v1]; *“It was not their fault”*[2095\_v3]), are unaware of the damage that asbestos might cause (17 cases) *“People were not warned of asbestos damage it could cause”* [2065\_v3]), or that business and government should take responsibility for people’s safety regarding asbestos (9 cases) *“It is the duty of business/ government to lay down guidelines concerning asbestos”* [1048\_v3]).

20 of the 23 respondents who switched to rating A (smoking cancer) as worse than B (asbestos cancer) in the ‘contextual’ version gave written comments. The most dominant theme to emerge here was that the blame issue may not be as clear-cut as it looks (6 cases - *“People chose to work in certain industries even when deaths related to asbestos were*

*known*” [1039\_v3]; “*People who started smoking in their adolescence knew as much about smoking related illnesses as people working in asbestos industry*” [2029\_v1]; “*Lung cancer caused by smoking is the fault of business/ government*” [1037\_v3]). All other types of reasons were only raised by one or two respondents and included: the difficulties associated with stopping smoking (“*It’s harder to quit smoking*” [1036\_v3]); being a smoker (“*A lapsed smoker*” [2159\_v1]); or the difficulties doing anything about asbestosis (“*Asbestos damage was done years ago and cannot be rectified now*” [2035\_v3]).

### 3.4.3 Work-related Cancer vs. Car Drivers

30 of the 48 respondents who switched to rating A (work-related cancer deaths) as worse than B (car driver deaths) in the ‘contextual’ version gave written comments. The majority of written comments were linked to the ‘blame’ and ‘quality of life’ attributes – i.e., the work-related cancer victims had no choice over the situation and their deaths were caused by others (14 cases “*The individuals had no control*” [2027\_v1]; “*Awful, why so many deaths, again due to business or government. In scenario B, the car drivers are to blame for the deaths*” [2113\_v2]), or that the work-related cancer victims had greater or longer suffering (10 cases “*I think A is much worse than B because suffering one-two years before death is more painful than die instantaneously*” [2045\_v2]).

More generally however, a number of respondents expressed the view that workers should not be put at risk in the work-place (10 cases “*You should not be put at risk in a work place and should be protected by law*” [2058\_v1]), or that these deaths could have been prevented (3 cases “*Cancer could have been prevented if they hadn’t been exposed to the chemicals at work due to the business*” [1087\_v2]).

No respondents switched to rating B (car driver deaths) as worse than B (work-related cancer deaths) in the ‘contextual’ version.

In summary, the qualitative data highlight two main reasons why the ‘contextual’ questions produce different responses to the ‘generic’ questions. The first reason is that the five attributes used to describe the causes of death in the ‘generic’ questions may not fully capture the differences between the deaths that participants want to take into account when giving their ratings. For example, in the comparison between car drivers and rail passengers some of the reasons given picked up on the fact that rail deaths were regarded as worse because participants expect the railways to be a safe place or that rail accidents are avoidable. On the other hand some reasons for regarding car accidents as worse than rail accidents picked up on considerations about the higher baseline risk. Deaths from rail accidents were not as bad as car accidents because rail accidents were much rarer. The other main issue that may not have been captured by the quality of life attribute was some participants’ perceptions of the particularly nasty nature of the rail accidents in particular.

The second reason is that the interpretation of the levels used to describe the attributes may cover quite a wide range of different situations and providing a context may make the distinctions between levels less clear-cut. This is most clearly illustrated by the ‘individuals themselves’ and ‘business or government’ levels of the ‘blame’ attribute both of which differentiated between the Car Accident vs. Rail Passenger and Smoking Cancer vs. Asbestos Cancer pairings. However, it is notable that whereas in the Car Accident vs. Rail Passenger pairing more of the reasons for rating the rail passenger deaths as worse than the car accident deaths were related to the fact that rail accident victims were not to blame, in the Smoking

Cancer vs. Asbestos Cancer pairing more of the reasons for rating the asbestos cancer deaths as worse than the smoking cancer death were related to the fact that smokers were to blame.

This suggests that the level ‘business or government’ was regarded as more deserving of a ‘worse than’ rating when it was used to describe rail passenger deaths than when it was used to describe asbestos cancer deaths, and similarly the level ‘individuals themselves’ was regarded as less deserving of a ‘worse than’ rating when it was used to describe smoking cancer deaths than car driver deaths. In other words although participants may have accepted that ‘business or government’ was most to blame in the cases of asbestos cancer deaths and rail passenger deaths, the actual degrees of business or government responsibility might be perceived quite differently by participants for these two types of deaths. Similarly participants might perceive the actual degrees of individual responsibility associated with car driver deaths and smoking cancer deaths quite differently. As shown by some of the reasons for choosing smoking cancer deaths over asbestos cancer deaths, when contextual information is provided people are able to draw upon their own knowledge about who is responsible such that the distinction between individuals and business/ government is less clear-cut than it might have been when the description was presented generically.

#### **4 DISCUSSION**

Our study sought to establish the extent to which ‘generic’ descriptions where causes of death are described merely in terms of their standing on a number of key attributes (age, severity and duration of pain and suffering, blameworthiness) are predictive of the response to ‘contextual’ descriptions where the causes of death are identified. We find evidence of differences in responses between the ‘generic’ and ‘contextual’ questions, indicating that it is

unlikely that a generic model of ‘bad deaths’ could ever be estimated in order to estimate the relative weight attached to preventing fatalities across different sectors of the economy.

The qualitative data suggest that two factors lead to differing perceptions between the ‘generic’ and ‘contextual’ questions. Firstly, some influential variable(s) were omitted from the ‘generic’ questions (for example, the ‘violent’ nature or ‘avoidability’ of deaths on the railways). It is possible that the use of a larger number of attributes may have overcome this problem but there is always a tension between the need to describe the totality of influencing factors and the need to arrive at a manageable set of attributes (Coast and Horrocks, 2007)

Secondly, when the causes of death were provided certain categories of included variables were interpreted somewhat differently (for example, greater importance seemingly being attached to the ‘blame’ dimension in some cases). Whilst the problem of omitted variables can be solved at least conceptually, there is no obvious means of overcoming the finding that the interpretation of attributes is context dependent. Thus, our study highlights the difficulty of trying to use a ‘generic’ set of attributes to anticipate people’s responses towards a set of attributes where the specific cause of death is identified. Similarly, Smith (2008) reported that attaching labels such as ‘stroke’ and ‘bowel cancer’ to health state descriptions led respondents to generate additional symptoms and to interpret the ‘prognosis’ information differently.

Detecting differences between labeled and generic descriptions does not, however, allow us to conclude that one set of values is superior to the other. We argued here that ‘sensitivity to theoretically relevant factors’ – such as the numbers dying – is one criterion against which the validity of responses may be measured (Loomes, 2006). Whilst there are other criteria

against which validity may be assessed (such as insensitivity to theoretically irrelevant factors- such as framing effects), we elected to focus on sensitivity to the numbers dying as that has become the ‘acid test’ in the WTP literature. Whilst we found differences in the sensitivity to the numbers of deaths between the ‘contextual’ and ‘generic’ questions, there was no clear pattern to this finding. Hence, it is difficult to point to any empirical support for the superiority of one set of responses over another using that particular criterion.

In the absence of any clear empirical evidence of the ‘superiority’ of one set of descriptions over another, we turn to more conceptual considerations of whether generic or contextual descriptions are appropriate for resource allocation decisions. Perhaps the most obvious case for the use of generic, rather than contextual, descriptions is that the former aid resource allocation decisions in allowing comparability across competing programmes. After all, if the *specific* cause of death (for example, as a driver in a car accident) did not matter to respondents, a ‘generic’ model – such as that outlined in the appendix could be used to predict the utility of life saving intervention with any particular combination of attribute levels. That is, decision makers could simply ‘plug in’ the relevant characteristics of any particular death in order to estimate the relative weight attached to preventing a fatality of that nature. As long as the deaths may be described in terms of the same attributes, this would achieve comparability in decision both within and across sectors responsible for health and safety.

Perhaps the most systematic use of ‘generic’ outcome measures is in the area of health. In economic evaluations of health care treatments and interventions health outcomes are generally valued in terms of Quality Adjusted Life Years (QALYs) derived from ‘generic’ health state classification systems such as the EQ 5D – a measure which values five dimensions of health-related quality of life: mobility, self-care, usual activities, pain, anxiety/



depression (Dolan et al., 1996). Indeed, the National Institute for Health and Clinical Excellence (NICE) *require* all health outcomes to be assessed using the EQ 5D in submissions to them (NICE, 2008). So, moving a patient from one EQ 5D health state to another is always assigned the same value, irrespective of how that change is brought about. In this way, the health benefits yielded by cancer therapies may be compared directly to those yielded by treatments for dementia, heart disease, arthritis or indeed anything else so long as outcomes are expressed in movements on the EQ 5D.

Another argument for the use of generic outcomes in health is they factor out ‘emotive’ views of, for example, conditions like cancer and put all conditions/patient groups on an equal footing in the competition for scarce resources. The issue of ‘emotive’ views can be considered in the framework set out by Loewenstein and O’Donoghue (2004) who make a distinction between ‘affective’ system and ‘deliberative’ system for decision making – where affective are based on emotive impressionistic reactions, and deliberative involve more systematic weighing up of consequences. For example, Slovic et al (2004, 2007) argue that, in responding to questions involving contexts that carry strong negative affect meanings, respondents are less sensitive to probability information than contexts that carry less affect. For example, with hazards like nuclear power and exposure to small amount of toxic chemicals, the negative consequences of these risks may make respondents more concerned and sensitive to the possibility of these risks, and less sensitive to information about the actual probability of these risks. It could be argued that, in such cases, the affective responses – based on emotive impressionistic considerations – are ‘distorting’ responses and ought to be ‘factored out’.

In addition, the evidence from the psychiatric literature cited in the introduction suggests that labels such as ‘schizophrenia’ have strong, negative effects on people’s perceptions and judgments about individuals with mental illness (Link et al., 1987). This may be due to the fact that the ‘label’ of schizophrenia evokes stereotypic perceptions that may be inaccurate. For example, people may believe that individuals with schizophrenia are violent more often than is truly the case. Assigning the label schizophrenic to an individual may lead others to interpret aspects of their behaviour in a way that is consistent with the negative stereotype. This suggests that ‘labelling’ the cause of death may induce a stereotype of the typical victim thereby biasing the judgment about the ‘badness’ of that death. For example, labelling a victim ‘a smoker’ may evoke a negative stereotype of a typical victim that shifts attention away from the unpleasant nature of the death that would otherwise have featured in peoples’ thinking.

However, whilst there may well be *normative* arguments against the use of contextual descriptions, the results of this research suggest that the public may well reject the policy implications that arise from the use of their own ‘generic’ valuations. So, even if we have reason to believe that public views of, for example, nuclear power and toxic chemicals are somehow ‘distorted’, their existence does create problems for decision makers. Similarly, if the public *does* have a strong aversion to particular diseases, then they may well reject NICE decision making that seeks to factor out such considerations. This problem may be exacerbated if public attention is focused on a single, identifiable victim as was the case with Herceptin for breast cancer.

One solution to the ‘problem’ of the gap between generic and contextual valuations may be to adopt more intensive, interactive valuation methodologies. In his review of the literature of

value elicitation in the fields of health, safety and the environment Loomes (2006) suggested that policy makers should first try to elicit from respondents a broad list of principles on which to allocate resources. The implications of these rules could then be shown to respondents to ensure that they agree with the implications. In the current context, the feedback mechanism would certainly involve a discussion of the context and example of the types of decisions the general rules would dictate. Were respondents to overwhelmingly reject the implications of general rules once the implications of the rules are made clear, the rules would need to be revised. This suggests an iterative, citizen's jury, type approach to valuation although it is an open question whether the gap between 'contextual' and 'generic' rules would necessarily be narrowed by such a process.

## 5 REFERENCES

- Beattie, J., Covey, J., Dolan, P., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N., Robinson, A., & Spencer, A. (1998). On the contingent valuation of safety and the safety of contingent valuation: Part 1 - 'Caveat Investigator'. *Journal of Risk and Uncertainty*, 17(1), 5-25.
- Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77-101.
- Carson RT (1997) Contingent valuation surveys and tests of insensitivity to scope. In Kopp RJ, Pemmerhene W and Schwartz N (eds) *Determining the Value of Non-Marketed Goods: Economic, Psychological and Policy Relevant Aspects of Contingent Valuation Methods*. Kluwer, Boston.
- Coast J, Horrocks S. (2007) Developing attributes and levels for discrete choice experiments using qualitative methods. *Journal of Health Services Research & Policy*, 12, 25-30.
- Covey J, Jones-Lee M, Loomes G, and Robinson A (1998) Valuing the Prevention Of Food-Borne Illness: Some Limitations Of Consumers' Willingness To Pay, *Risk Decision and Policy*, 3 (3): 245-259.
- Desaignes B and Rabl A (1995) Reference values for human life. In Schwab-Christie NG and Soguel NC (eds), *Contingent Valuation, Transport Safety and the Value of Life*. Boston Kluwer.
- Dolan, P., Gudex, C., Kind, P. & Williams, A. (1996). The time trade-off method: results from a general population study. *Health Economics*, 5, 141-154.
- Dubourg WR, Jones-Lee MW and Loomes G (1997) Imprecise preferences and survey design in contingent valuation. *Economica* 64: 681-702.
- Gerard, K., Dobson, M., and Hall, K. (1993) Framing and labeling effects in health descriptions: Quality adjusted life years for treatment of breast cancer. *Journal of Clinical Epidemiology*, 46, 77-84.
- Fryer, J.H & Cohen, L. (1988). Effects of labeling patients “psychiatric” or “medical”: Favorability of traits ascribed by hospital staff. *Psychological Reports*, 62, 779-93.

- Hammitt JK and Graham JD (1999) Willingness to pay for health protection: inadequate sensitivity to probability? *Journal of Risk and Uncertainty* 18: 33-62.
- Jenni, K.E. & Loewenstein, G. (1997). Explaining the “Identifiable victim effect”. *Journal of Risk and Uncertainty*, 14, 235-57.
- Jones-Lee, M.W., Hammerton, M., Philips, P.R., (1985) The value of safety – results from a national sample survey. *Economic Journal*, 95(377), 49-72.
- Jones-Lee MW, Loomes G and Philips PR (1995) Valuing the prevention of non-fatal road injuries: contingent valuation vs standard gambles. *Oxford Economic Papers* 47: 676-695.
- Kogut, T. & Ritov, I. (2005). The “Identified Victim” effect: An identified group, or just a single individual? *Journal of Behavioral Decision Making*, 18, 157-67.
- Krippendorff, K. (1980). *Content Analysis: An Introduction to Its Methodology*. Newbury Park, CA: Sage.
- Link, B.G., Cullen, F.T., Frank, J. & Wozniak, J.F. (1989). The social rejection of formal mental patients: Understanding why labels matter. *American Journal of Sociology*, 92, 1461-1500.
- Loewenstein, G., & O’Donoghue, T. (2004), Animal spirits: Affective and deliberative processes in economic behaviour. Cornell University, Center for Analytic Economics Working Paper #04-14. Available online at <http://www.arts.cornell.edu/econ/CAE/04-14.pdf>
- Loman, A.L. & Larkin, W.E. (1976). Rejection of the mentally ill: An experiment in labeling. *Sociological Quarterly*, 17, 555-60.
- Loomes, G. (2006) (How) Can we value health, safety and the environment? *Journal of Economic Psychology*, 27(6), 713-736.
- Louviere, J.J., Hensher, D.A., & Swait, J.D. (2000). *Stated Choice Methods: Analysis and Application*. Cambridge: Cambridge University Press.
- NICE - National Institute for Health and Clinical Excellence (2008). *Social Value Judgments: Principles for the Development of NICE Guidance*. Second Edition. Available online at <http://www.nice.org.uk/media/C18/30/SVJ2PUBLICATION2008.pdf>
- Rabin, R., Rosser. R.M., & Butler, C. (1993) Impact of diagnosis on utilities assigned to states of illness. *Journal of the Royal Society of Medicine*, 86, 444-448.

- Robinson, S., Bryan, S., (2001) 'Naming and framing': an investigation of the effect of disease labels on health state valuations, Paper presented at the Health Economists' Study Group Meeting, January, University of Oxford.
- Sackett, D.L., Torrance, G.W.(1978) Utility of different health states as perceived by general public, *Journal of Chronic Diseases*, 31(11),697-704.
- 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.
- Slovic, P. (1992). Perceptions of risk: Reflections on the psychometric paradigm. In S. Krimsky, and D. Golding. (Eds). *Social Theories of Risk*. Westport, CT: Praeger, pp. 117-152.
- Slovic, P., Finucane, M.L., Peters, E., MacGregor. D.G. (2004) Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis*, 24(2), 311-322.
- Slovic, P., Finucane, M.L., Peters, E., MacGregor, D.G. (2007) The affect heuristic *European Journal of Operational Research*, 177(3), 1333-1352.
- Small, D.,A. & Loewenstein, G. (2003). Helping a victim or helping the victim: Altruism and identifiability. *Journal of Risk and Uncertainty*, 26(1), 5-16.
- Smith, R, (2008), Contingent Valuation in Health Care: Does it Matter How the 'Good' is Described, *Health Economics*, 17: 607-617
- 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.
- Yeung RYT, Smith RD McGhee SM (2003) .Willingness to pay and size of health benefit: an integrated model to test for 'sensitivity to scale', *Health Economics*, 12 (9): 9-14.

## 6 ACKNOWLEDGEMENTS

The research reported in this article was carried out under a project funded by the Health and Safety Executive. However, the opinions expressed in this article are solely the responsibility of the authors and do not necessarily reflect the views of the project sponsor. The authors would like to thank Aki Tsuchiya for discussing an earlier version of the paper at a Health Economics Study Group Meeting.

## Appendix: The generic DCE model

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. Standard utility theory entails that respondents ought to be multiplying the disutility of a particular type of death by the number of those deaths. We therefore estimated a model that was multiplicative in the number of deaths as follows:

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

where  $P(B)$  is the probability that a respondent will consider scenario B to be worse than A,  $N_A^\alpha$  and  $N_B^\alpha$  are the number of deaths in scenarios A and B respectively raised to power  $\alpha$ , and  $\gamma_i$  is the coefficient on the  $i^{\text{th}}$  attribute. When  $\alpha$  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  $\alpha$  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  $\alpha$  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 type’ of death was accorded a value of 1. We first estimated the model with  $\alpha = 1$ , but a number of the coefficients appeared to have the wrong sign and did not appear to fit the data at all well. We then explored other values of  $\alpha$ . A grid-search showed that the log-likelihood function was minimized (i.e. the multiplicative model fitted best) when  $\alpha = 0.2$ . The results of estimating the model with  $\alpha = 0.2$  are shown in Table 3. Observations were not independent (as each respondent contributed 12 observations), so standard errors were adjusted to allow for clustering by respondent.



**Table 3: The multiplicative ‘generic’ DCE model**

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		

The ‘base type’ of death used in the model is: age (over 60s), severity (bit worse than normal), duration (last few minutes), blame (nobody in particular.). The *deaths* variable is ‘offset’ to make the disutility of the base case type of death equal to 1. The remaining variables are prefixed by a ‘D\*’ to indicate that they are multiplicative in the number of deaths. 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.

### Quantifying the ‘badness’ of deaths

The DCE model provides a ‘generic’ tool that may be used to estimate the relative badness or disutility of various different types of deaths that may be described in terms of levels on the four attributes. If such a generic model were appropriate, 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 death to take on a value of 1. So, *if the actual cause of death* did not matter, we could simply describe deaths in terms of the attributes in the model and calculate the relative badness of the two types of deaths. A couple of examples are illustrated below.

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 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 DCE model we would get:

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

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

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:

$$\frac{U_{deathA}}{U_{deathB}} = \frac{2.144}{1.348} = .59$$

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 above, this example would then yield a ratio of 1.107:

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

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

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

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.

The results in the body of this paper, however, raise doubts about whether preferences over bad deaths may be adequately captured in a ‘generic’ model such as this as context does appear to matter.