Incorporating psycho-social considerations into health valuation: an experimental study

Richard Cookson

LSE Health, LSE Health, London School of Economics, Houghton Street, London WC2A 2AE, UK

Received 1 December 1997; received in revised form 3 November 1998; accepted 4 June 1999

Abstract

This study used multiple elicitation question modes and group discussions to value psycho-social considerations associated with public perceptions of health risks. Values for saving 100 statistical lives in six different policy contexts varied by ±50%, largely due to psycho-social considerations of choice and control. A ‘currency effect’ was found: value of life differentials were smaller when questions were framed directly in the currency of lives saved rather than willingness to pay or public spending to save lives. Discussions reduced but did not eliminate inconsistencies, and slightly increased the relative valuations of health interventions involving less choice and control by those at risk. © 2000 Elsevier Science B.V. All rights reserved.

JEL classification: J17

Keywords: Value of life; Economic evaluation methodology; Risk perception; Contingent valuation; Focus group

1. Introduction

When evaluating the benefits of health interventions, economists base their calculations on scientific estimates of the expected effects on life and health. However, these calculations typically do not incorporate psycho-social factors of the kind that psychologists have shown to influence non-scientific lay perceptions of risk, such as the extent to which risk is involuntary, unknown or potentially involuntarily.
To the extent that psycho-social factors influence people’s health (for instance, their level of anxiety) or their well-being more broadly (for instance, their freedom to choose health risks), it seems reasonable to attempt to quantify the benefits that flow from them.

The small-scale experimental study reported in this paper explores one way of quantifying the benefits that flow from psycho-social factors. It does so using multiple “relative valuation” elicitation question modes in conjunction with group discussions. Relative valuation questions value similar items relative to one another in terms of a suitable unit of health (in this study, statistical lives). This unit of health can then be given a monetary price per unit. Group discussions allow respondents to engage in more intensive deliberation than is possible using conventional questionnaires or interviews, so that respondents can arrive at more “considered” preferences.

Six areas of health and safety policy are investigated, which may be expected to differ in terms of psycho-social characteristics: air pollution, birth control pills, car accidents, food poisoning, medical radiation and rail accidents. Four different elicitation question modes are used: Priorities, Matching, Willingness to Pay and Relative Monetary Valuation (see Section 4). The study focuses on fatalities, rather than illnesses and injuries; and on preventative measures, rather than treatments. In principle, however, the methods developed in this study could be used in all areas of health economics — and, indeed, in all areas of economic evaluation which involve a gap between public perceptions and scientific assessments.

This paper is structured as follows. Section 2 explains why this study uses “relative valuation” elicitation questions. Section 3 then explains why it uses a “deliberative” group discussion approach. Section 4 gives details of the design of the study. Section 5 reports the relative values of life estimated in the study. Section 6 reports the qualitative reasons that respondents gave for their valuations. Section 7 examines the effects of elicitation question mode on responses. Section 8 examines the effects of deliberation and discussion on responses. Section 9 discusses the results and draws out the methodological lessons for researchers engaged in eliciting individual preferences for health and safety. Finally, Section 10 concludes with a brief summary of the main findings and implications.

2. The relative valuation approach

Health economists have had great difficulty valuing psycho-social considerations associated with public perceptions of risk and the wider policy context. This

---

1 Psychometric research into risk perception is reviewed in Slovic (1992). The standard approach is to ask subjects a series of questions in which they rate several hazards according to perceived risk and then according to various psycho-social dimensions of risk.
problem of psycho-social considerations in health evaluation is analogous to the problem of “existence values” in environmental valuation, which refer to benefits the individual obtains from the very existence of environmental or other resources that she cares about but does not physically use. Psycho-social factors and existence values both seem worthy of inclusion into economic evaluation — yet both are extremely difficult to measure.

The standard QALY approach to health economic evaluation does not try to incorporate psycho-social factors. Instead, risk is typically dealt with by calculating the expected value of health benefits: the value of each possible health outcome is multiplied by a scientific point estimate of the corresponding probability of occurrence. Although questionnaire methods for valuing health states do incorporate individual perceptions of the individual’s own (context-free) health status, they are not designed to incorporate the psycho-social factors that enter into individual perceptions of risk and the wider policy context.²

In principle, the QALY calculus could be extended to examine how psycho-social factors influence people’s health, in terms of levels of anxiety and depression. A fine-grained measure of anxiety levels could be incorporated into a generic health status index, and a way found to measure changes in levels of anxiety associated with different health interventions. A potential practical difficulty with extending health state measurement tools in this way is that respondents might find it hard to compare small changes in health (ex ante anxieties about health risks) against large ones (ex post serious diseases and injuries). Perhaps more fundamentally, however, this procedure is open to the objection that people may be concerned with psycho-social considerations not only as a determinant of ill-health or “health-related quality of life”) but also as an influence on wider aspects of individual well-being.

This suggests that psycho-social considerations may require a “willingness to pay” approach to valuation, which can potentially take into account wider aspects of individual preferences and well-being. However, willingness to pay methodologies — in particular, the “revealed preference” approach and the “contingent valuation” approach — have run into well-known practical difficulties.

The “revealed preference” approach uses micro-econometric models to extrapolate valuations from market data. However, suitable market data are unavailable for most of the areas of health and safety that interest policymakers, because the various “markets” for health are incomplete and heavily regulated. With the notable exception of occupational safety, researchers have had great difficulty

² The health state measurement method known as “Standard Gamble” may be influenced by risk perceptions to some extent, since it involves choices among artificially constructed gambles concerning health risks (Torrance, 1986). However, it does not take account of the policy context, nor of perceptions of very small risks associated with some policy decisions. Moreover, it can be argued that the value set on a particular health state under certainty should not be influenced by risk attitudes anyway (Broome, 1993).
identifying contexts in which the available market data are sufficiently detailed to disentangle the factors other than health and safety that may be influencing behaviour.

The "contingent valuation" approach essentially uses a questionnaire to ask individuals how much they are willing to pay for a specified benefit (or, less often, willing to accept in compensation for a specified loss). However, concerns have been raised that contingent valuation may be vulnerable to a serious survey design bias known variously as "part-whole bias", "embedding effects" and "scope effects" (Fischhoff et al., 1993; Diamond and Hausman, 1994; Baron and Greene, 1996; Bateman et al., 1997). In general terms, this potential bias means that respondents may state much the same valuation for a larger or more comprehensive bundle of benefits as for a smaller bundle. In the specific context of valuing health risks, it means that respondents may place a similar value on larger changes in health risk as they do on smaller changes.

A team of UK-based researchers, who have undertaken a number of studies on the contingent valuation of health risks, have recently reported evidence of extensive and persistent insensitivity to the size of the reductions in health risk (Beattie et al., 1998b). They have also reported evidence that contingent valuation responses tend to be overly sensitive to irrelevant "cues", such as payment cards (Dubourg et al., 1998). A recent survey of the field found 700-fold variations in estimated monetary values of life between different contingent valuation studies, and concluded that much of this variation can be attributed to measurement error (Beattie et al., 1998a). More specifically, the survey found that studies that examined smaller changes in risk tended to obtain higher estimates of the value of saving a single life, a pattern consistent with a widespread and serious problem of scope effects. However, the reliability of contingent valuation is a controversial topic, and other researchers have argued that problems of this kind are due to inadequate survey design (Hanemann, 1994, 1995; Carson and Mitchell, 1995).

Due to these various limitations of standard valuation approaches, this study adopts a "relative valuation" approach to valuing psycho-social considerations (Jones-Lee and Loomes, 1995). Rather than trade off money against health risk, as in contingent valuation, respondents can be asked to trade off one health risk against another, in terms of some common unit of health (such as statistical lives, or QALYs). In effect, this means that psycho-social considerations are valued relative to the common unit of health. The resulting relative valuations can then be converted into "absolute" monetary valuations, for use in cost-benefit analysis, by setting a monetary value on the common unit of health. This method has been used

3 They report that Covey et al. (1995) estimated a food poisoning value of life of £48,840,000 in 1995 prices using an exceptionally small risk reduction of 1.67 in 10 million, and a study by Acton in 1973 estimated a heart attacks value of life of £69,800 in 1994 prices using an exceptionally large risk reduction of 2 in 10000.
to compare underground rail transport safety with road safety, and has found evidence for a substantial premium on the former (Jones-Lee and Loomes, 1995). A similar relative valuation strategy has been employed to derive valuations for preventing non-fatal casualties relative to fatalities (Viscusi et al., 1991; Jones-Lee et al., 1995).

3. The deliberative approach

In addition to the use of relative valuation elicitation questions, the present study departs from standard questionnaire approaches by using group discussions. It employs group and pair discussions to encourage deliberation by respondents, allowing them to develop their views and to express more ‘‘considered’’ preferences. This approach is more expensive and more demanding of both respondents and researchers than conventional questionnaire methods, since respondents spend longer deliberating about their answers, and researchers spend longer analysing the results.

A handful of studies on the interface between economics and psychology have examined the issue of psycho-social factors using non-deliberative questionnaire methods (i.e., without group discussions). These studies asked several ‘‘willingness to pay’’ (WTP) questions across a range of different life saving interventions to the same sample of people (Mendeloff and Kaplan, 1989; McDaniels et al., 1992; Savage, 1993 and Gregory and Lichtenstein, 1994). The valuation questions were then followed by a series of rating questions concerning the risk perception factors that psychologists use to explain people’s perceptions of risk.

The three primary factors are ‘‘Dread’’, ‘‘Unknown’’ and ‘‘Number’’. The ‘‘Dread’’ factor refers to the degree of unease that people feel about a particular risk, and is typically positively correlated with secondary factors such as lack of control over risk, the involuntary nature of risk, and potentially catastrophic consequences. The ‘‘Unknown’’ factor refers to how far risks are well-defined and quantifiable, and is typically positively correlated with secondary factors such as how far the risk is known to science or to those exposed, the observability of the risk, and how new the risk is. Finally, the ‘‘Numbers’’ factor relates to the perceived number of people exposed to the risk. These studies found up to 6-fold variations in values for life saving between different contexts, which were strongly correlated with the three primary risk perception factors.

However, the ‘‘unconsidered’’ risk preferences elicited by conventional surveys of this kind may not be suitable data for the purposes of economic evaluation, since they may reflect misperceptions of risk that the individual herself would wish to correct on further reflection. More generally, unconsidered preferences may be influenced by ‘‘framing effects’’ and other psychological effects and biases (Payne et al., 1992; Dolan, 1997). The aim of the deliberative approach is to
provide respondents with an opportunity to iron out such errors, biases and inconsistencies, so as to arrive at "true" preferences which accurately reflect their real interests or well-being.

One objection to deliberative methods is that deliberation and reflection may fail to eliminate biases and errors. Worse, deliberation in group settings may actually increase biases and errors, due to herd behaviour or to dominant personalities within the group. In other words, there is no guarantee that "considered" preferences are any better than "unconsidered" preferences.

However, we can identify two independent criteria for judging whether or not considered preferences are "true" preferences. First, the internal consistency of responses, and second, the plausibility of the qualitative reasons that people give for those responses. The criterion of internal consistency is well accepted in the economics literature (e.g., Arrow et al., 1993), although the criterion of plausible underlying reasons is controversial and many economists would reject this as a way of assessing evidence on individual preferences.

At a more fundamental level, a number of psychologists and behavioural decision theorists have argued that "true" preferences do not exist, because preferences are "constructed" in response to stimuli rather than revealed or discovered (Slovic, 1995). However, the implications of this "constructed preferences" view for economic evaluation are open to interpretation and have yet to be fully developed. One possible implication is that all methods of economic evaluation are equally (in)valid, and give different results because they construct preferences in different ways. On the other hand, it might be argued that some ways of constructing preferences for use in economic evaluation are inherently better than others, even though the resulting preference cannot be judged "true" or "false" using independent criteria such as consistency and plausibility. If so, deliberative methods of preference elicitation may be a better way of constructing preferences than non-deliberative methods (Gregory et al., 1993).

A third criticism of the deliberative approach is that it is too paternalistic and authoritarian, and fails to respect people's actual, unconsidered preferences. However, as Goodin (1995) argues, attempting to improve or "launder" individual preferences may be in line with the individual's own wishes. The individual may recognise that there are circumstances under which he would wish his own preferences to be "laundered" for valuation purposes. Given the unfamiliar consequences, great uncertainties and long delays involved, it is reasonable to suppose that one's own "unconsidered" preferences between health risks will be biased by imperfect information, errors and misunderstandings.

A fourth criticism is that the deliberative approach places too much emphasis on the subjective judgement of the investigator in setting the agenda for deliberation and discussion. To put it crudely, the investigator might brow-beat respondents until they accept his point of view. In reply, however, it can be pointed out that the way a question is framed in traditional questionnaire methods can have just as much influence on results as the way the discussion agenda is set in
deliberative methods. The important thing is that full details of the survey design are reported, so that all potential forms of survey design bias may be assessed.

4. Design of the study

The study investigated the values for saving life in different health and safety contexts held by a particular group of subjects. The aim was to examine how far relative valuation questions in conjunction with a deliberative approach can provide a valid measure of true economic preferences for these particular subjects, based on consistency checks and qualitative evidence. No claim is made that the results are representative of a wider population suitable for policymaking purposes.

The sample consisted of 52 parents from a local primary school, who were mostly women of child-bearing age. Each subject was paid £15, with a further £5 each being donated to the school fund. The design was pre-tested by two pilot studies, also with paid members of the public, in December 1995 and February 1996; the main study was conducted in May 1996.

The main study consisted of 10 workshops, with four to six participants in each, which involved both tape-recorded discussions and responses to value elicitation questions. Each workshop lasted about 2 h. It was felt that six policy contexts was about as many as subjects could comfortably handle in one sitting. To test whether the number of items being valued made any difference, half the sample were given questionnaires with only four contexts. The contexts were as follows: air pollution (A), birth control pills (B), car accidents (C), food poisoning (F), medical radiation (M), and rail accidents (R). They were chosen on the basis that (a) they would be reasonably familiar to most participants, (b) they would allow credible public risk reduction scenarios, and (c) they would allow a degree of comparability with previous studies.

It was decided to use public policy scenarios, rather than consumer purchase scenarios, in order to encourage respondents to come to an appropriate understanding of the policy situation in which their expressed valuations might ultimately be used. Furthermore, by considering public health risk trade-offs in this way, respondents are able to take account of all the relevant factors and not just those relating to their own private level of health risk. The scenarios focused upon fatalities, rather than non-fatal detriment to health, in order to keep the questions as straightforward as possible. Each health intervention was briefly described to respondents on “information cards”, which are reproduced in Appendix A.

These valuation scenarios were all carefully researched and pre-tested for clarity and credibility. Since there were six scenarios, rather than one, they were briefer than those in most contingent valuation surveys. It can be argued that this brevity is an advantage, given the psychological fact that people tend to simplify their cognitive tasks when presented with detailed information (Slovic, 1995).
Hence, the inclusion of too much detail may distract attention away from essential value-bearing characteristics and towards irrelevant “cues”, such as the perceived costs of the programmes, or associated non-health benefits (e.g., reduced traffic congestion).

In order to elicit relative valuations across the different policy areas, the magnitude of public risk reduction was held constant across each health intervention, at an expected 100 fewer deaths in the UK over the next 10 years. This way of communicating risk information differs from the traditional approach employed in willingness to pay studies, which is to present information on the changed individual probability of dying. However, given the focus on public safety, and the well-known problems people have dealing with probability concepts, it was felt that this more direct approach using numbers of deaths would be easier for subjects to handle (a similar approach is adopted in Beattie et al., 1998b).

Four different value elicitation question modes were used. The first value elicitation question was a “Priorities” question, which asked respondents to place each safety improvement in ordinal rank order of priority, allowing ties. The Priorities question is reproduced in Fig. 1.

The second form of value elicitation question was a cardinal relative valuation question which compares two policy contexts relative to one another, and is known as a “Matching” question (Jones-Lee and Loomes, 1995) or a “Person-Trade-Off” question (Nord, 1995). The first part of the Matching question used in this study is a pairwise choice question, an example of which is shown in Fig. 2.

Those who state no strict preference between saving the same number of lives in the two contexts are assumed to value life saving in both contexts equally. However, those who state a strict preference for one of the two contexts then proceed to one of two follow-up questions. These ask the respondent to mark on a numerical scale how many deaths prevented in their preferred context they would regard “equally as good as” a baseline of 100 deaths prevented in the other context. An example follow-up question is shown in Fig. 3.

Respondents were given three Matching questions: R vs. C, C vs. F and then F vs. R. Due to time constraints, contexts A, B and M were not valued using Matching questions.

The third relative valuation question was a “Willingness to Pay” (WTP) question, shown in Fig. 4, which asked for valuations of all six contexts. This was an open-ended question that asked respondents how much their household would be prepared to pay each year in extra taxes over the next 10 years for each particular safety improvement. To obtain relative valuations, free from possible bias due to income and substitution effects, subjects were asked to answer each WTP question in isolation from the other WTP questions — that is, on the assumption that the other payments and safety improvements would not take place.

The fourth relative valuation question is a new type of elicitation question, developed for the purpose of this study, that can value several contexts relative to one another. This is the “Relative Monetary Values” question, shown in Fig. 5. It
Please place these public safety improvements in order of priority, starting with the improvement(s) that you think should have top priority.

Please assume that each safety improvement would **definitely** prevent a total of 100 deaths in the UK over the next ten years and would have **no other significant effects**.

Please assume that each improvement would cost the same amount for the UK taxpayer.

You may give more than one improvement the same priority number.

If you think they should all have the same priority, give them all priority number 1.

<table>
<thead>
<tr>
<th>PUBLIC SAFETY IMPROVEMENT OVER THE NEXT TEN YEARS</th>
<th>PRIORITY (1 = top priority)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMPROVEMENT C</td>
<td></td>
</tr>
<tr>
<td>100 fewer deaths from <strong>CAR ACCIDENTS</strong></td>
<td></td>
</tr>
<tr>
<td>IMPROVEMENT F</td>
<td></td>
</tr>
<tr>
<td>100 fewer deaths from <strong>FOOD POISONING</strong></td>
<td></td>
</tr>
<tr>
<td>IMPROVEMENT B</td>
<td></td>
</tr>
<tr>
<td>100 fewer deaths from <strong>BIRTH CONTROL PILLS</strong></td>
<td></td>
</tr>
<tr>
<td>IMPROVEMENT M</td>
<td></td>
</tr>
<tr>
<td>100 fewer deaths from <strong>MEDICAL RADIATION</strong></td>
<td></td>
</tr>
<tr>
<td>IMPROVEMENT A</td>
<td></td>
</tr>
<tr>
<td>100 fewer deaths from <strong>AIR POLLUTION</strong></td>
<td></td>
</tr>
<tr>
<td>IMPROVEMENT R</td>
<td></td>
</tr>
<tr>
<td>100 fewer deaths from <strong>RAIL ACCIDENTS</strong></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. Priorities question.

was an open-ended question that asked respondents to imagine that safety improvement **C** has been given a value of £100 m, and then asked them to place monetary values on the other safety improvements relative to this baseline value. Respondents were told that the figure of £100 m was a purely arbitrary number, and that it had been chosen for convenience to allow respondents to express the
Imagine that UK taxpayer’s money is available to spend on one of two possible public safety improvements.

Please assume that each safety improvement would definitely prevent a total of 100 deaths in the UK over the next ten years and would have no other significant effects.

HOW WOULD YOU DECIDE TO SPEND THE MONEY?

<table>
<thead>
<tr>
<th>IMPROVEMENT R</th>
<th>IMPROVEMENT C</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 fewer deaths from RAIL ACCIDENTS</td>
<td>100 fewer deaths from CAR ACCIDENTS</td>
</tr>
</tbody>
</table>

Please circle one box or put an = between them if you cannot decide.

Fig. 2. Preliminary choice part of the Matching question (an example).

relative importance they placed on the different safety improvements. It was also emphasised that the figure had nothing to do with government policy. Respondents were generally comfortable with this idea, and asked fewer clarificatory questions about this elicitation mode than about the ‘‘Willingness to Pay’’ elicitation question.

Three other elicitation question modes for eliciting relative valuations across several contexts at once were pre-tested in small pilot studies, but it was found that respondents were most comfortable with the Relative Monetary Values question. Comparisons in terms of units of ‘‘social loss’’ (Slovic et al., 1979) provoked many clarificatory questions and complaints that it was too abstract, as did Multiple Matching comparisons of lives saved in several contexts relative to 100 lives saved in a baseline context. Finally, division of a fixed budget among several items to yield ‘‘relative WTP’’ (Savage, 1993) caused difficulties of mental arithmetic when there were more than two or three items to compare.

It is worth noting that the Willingness to Pay question incorporates a notion of individual sacrifice, in terms of household income, whereas the other three questions are based on a notion of group sacrifice, in terms of taxpayer’s money or statistical lives lost to the community. The survey design used in this study allows direct experimental comparison between these different modes of elicitation.

The format of each workshop was as follows. They began with a 20-min group discussion, moderated by myself, which was structured on the basis of a

---

\*4 The choice of moderator is difficult for group discussions that form part of a health economic valuation study. Arguably, a professional health economist who lacks specialist training in interview skills may be able to handle discussions of this kind better than a professional interviewer who lacks specialist training in health economics.
Would you still prefer IMPROVEMENT R if it was expected to prevent fewer than 100 deaths?

Please put a ✓ next to each case in which you would still prefer IMPROVEMENT R.

Please put an X next to each case in which you would now prefer IMPROVEMENT C.

<table>
<thead>
<tr>
<th>Improvement C</th>
<th>Improvement R</th>
<th>Car Accident Deaths Prevented</th>
<th>Rail Accident Deaths Prevented</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>95</td>
<td>100</td>
<td>more than 95</td>
</tr>
<tr>
<td>100</td>
<td>90</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>85</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>80</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>75</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>70</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>65</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>55</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>45</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>40</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>35</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>30</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>15</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>fewer than 5</td>
<td>fewer than 5</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. Follow-up part of the Matching question (an example).
How much would your household be prepared to pay each year in extra taxes over the next ten years for each of these public safety improvements?

Please take each safety improvement in isolation and, in each case, assume that the other public safety improvements and extra taxes will not occur.

Please assume that that each safety improvement would definitely prevent a total of 100 deaths in the UK over the next ten years and would have no other significant effects.

Please bear in mind all of the other things that your household could do with the money.

<table>
<thead>
<tr>
<th>PUBLIC SAFETY IMPROVEMENT OVER THE NEXT TEN YEARS</th>
<th>HOUSEHOLD PAYMENT EACH YEAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMPROVEMENT C</td>
<td>£</td>
</tr>
<tr>
<td>100 fewer deaths from CAR ACCIDENTS</td>
<td></td>
</tr>
<tr>
<td>IMPROVEMENT F</td>
<td></td>
</tr>
<tr>
<td>100 fewer deaths from FOOD POISONING</td>
<td></td>
</tr>
<tr>
<td>IMPROVEMENT B</td>
<td></td>
</tr>
<tr>
<td>100 fewer deaths from BIRTH CONTROL PILLS</td>
<td></td>
</tr>
<tr>
<td>IMPROVEMENT M</td>
<td></td>
</tr>
<tr>
<td>100 fewer deaths from MEDICAL RADIATION</td>
<td></td>
</tr>
<tr>
<td>IMPROVEMENT A</td>
<td></td>
</tr>
<tr>
<td>100 fewer deaths from AIR POLLUTION</td>
<td></td>
</tr>
<tr>
<td>IMPROVEMENT R</td>
<td></td>
</tr>
<tr>
<td>100 fewer deaths from RAIL ACCIDENTS</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4. Willingness to pay question.

pre-established verbal protocol. The purpose of this discussion was to explain the purpose of the study as carefully as possible, and to introduce key economic concepts. The discussions began by familiarising subjects with the idea of using economic values for life saving and to introduce them to one of the relative valuation questions (a pairwise choice question which makes up the first part of a Matching question). During piloting, it was found that a “warm up” discussion of
To keep the numbers as simple as possible, imagine that the monetary value for preventing a "statistical" death from car accidents has been set at £1 million. This would mean that IMPROVEMENT C, which prevents a total of 100 deaths from car accidents, would be valued at £100 million.

Please set monetary values on each of the other safety improvements relative to this hypothetical monetary value of £100 million for IMPROVEMENT C.

Please assume that that each safety improvement would definitely prevent a total of 100 deaths in the UK over the next ten years and would have no other significant effects.

<table>
<thead>
<tr>
<th>PUBLIC SAFETY IMPROVEMENT OVER THE NEXT TEN YEARS</th>
<th>MONETARY VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMPROVEMENT C 100 fewer deaths from CAR ACCIDENTS</td>
<td>£ 100 million</td>
</tr>
<tr>
<td>IMPROVEMENT F 100 fewer deaths from FOOD POISONING</td>
<td>£ million</td>
</tr>
<tr>
<td>IMPROVEMENT B 100 fewer deaths from BIRTH CONTROL PILLS</td>
<td>£ million</td>
</tr>
<tr>
<td>IMPROVEMENT M 100 fewer deaths from MEDICAL RADIATION</td>
<td>£ million</td>
</tr>
<tr>
<td>IMPROVEMENT A 100 fewer deaths from AIR POLLUTION</td>
<td>£ million</td>
</tr>
<tr>
<td>IMPROVEMENT R 100 fewer deaths from RAIL ACCIDENTS</td>
<td>£ million</td>
</tr>
</tbody>
</table>

Fig. 5. Relative monetary values question.

this kind greatly reduced the number of clarificatory questions asked by respondents and virtually eliminated so-called “protest” responses. Protest responses are refusals to give an answer, or answers of zero, which are stated on grounds of disagreement with the questions (Mitchell and Carson, 1989). As a partial check on whether or not the warm up discussion influenced people’s responses, two subsamples were introduced to different valuation questions, involving a different pair of safety areas.
Following the warm up discussion, subjects had about 35 min to give initial answers to Priorities, Matching and Willingness to Pay questions. The questionnaire preamble, which lists a series of preliminary instructions and points to bear in mind, was read aloud and is reproduced in Appendix B. Each value elicitation question was read aloud, and was answered on an individual, rather than group basis. Subjects were invited to raise any points of clarification but asked not to discuss things at this stage.

After a short break, the group was split up into pairs and invited to give further consideration to the questions. During piloting, it was found that pair discussions ensured that all subjects took active part in the discussions, whereas discussions with three people allowed one of the subjects to remain silent. Two pairs took part in unmoderated discussions, with the task of agreeing upon a joint response to the Relative Monetary Values question. Any remaining subjects took part in a “feedback” interview with the moderator. These discussions and interviews lasted about 20 min, towards the end of which the subjects were asked to answer a set of attitude questions. The purpose of these discussions and interviews was to encourage subjects to articulate the thinking behind their responses and to give them the opportunity to re-consider their responses in the light of debate and discussion. Subjects were then brought back to the main table and asked to give final answers to all four type of value elicitation question on a fresh questionnaire — again, on an individual, rather than group basis.

5. Estimated relative values of life

Table 1 presents the final aggregate relative values for life saving held by the sample population in the six policy areas. Values are computed relative to area C (car accidents). Given the methodological purpose of this study, it was felt unnecessary to collect data on the demographic characteristics of the sample population or to analyse the statistical reliability of the mean response. An analysis

<table>
<thead>
<tr>
<th></th>
<th>$F$ ($n = 52$)</th>
<th>$B$ ($n = 27$)</th>
<th>$M$ ($n = 27$)</th>
<th>$A$ ($n = 52$)</th>
<th>$R$ ($n = 52$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Willingness to pay ($N = 51$)</td>
<td>0.98</td>
<td>0.44</td>
<td>0.63</td>
<td>1.34</td>
<td>0.85</td>
</tr>
<tr>
<td>Monetary values</td>
<td>0.89</td>
<td>0.57</td>
<td>0.69</td>
<td>1.81</td>
<td>0.84</td>
</tr>
<tr>
<td>Matching $^{a,c}$</td>
<td>1.19</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1.20</td>
</tr>
</tbody>
</table>

$^a$One respondent returned missing values for their four final WTP answers.

$^b$The Matching results are inflated by three outliers which appear to represent “protest” responses — that is, refusals to make explicit trade offs between lives to be saved, in the way suggested by the question.

$^c$Subjects were not asked to value areas $B$, $M$ and $A$ using matching questions.
Table 2
Household willingness to pay (each year for 10 years)

<table>
<thead>
<tr>
<th></th>
<th>C (n = 51)</th>
<th>F (n = 51)</th>
<th>B (n = 27)</th>
<th>M (n = 27)</th>
<th>A (n = 51)</th>
<th>R (n = 51)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arithmetic mean</td>
<td>£24.88</td>
<td>£31.81</td>
<td>£13.11</td>
<td>£14.05</td>
<td>£43.74</td>
<td>£23.31</td>
</tr>
<tr>
<td>Median and mode</td>
<td>£10.00</td>
<td>£10.00</td>
<td>£5.00</td>
<td>£10.00</td>
<td>£10.00</td>
<td>£10.00</td>
</tr>
</tbody>
</table>

of statistical reliability in relation to a wider population would of course be necessary for a full-scale study aimed at policy guidance, rather than methods development.

In broad terms, subjects valued statistical lives about the same in the context of food, rail and road safety; 50% higher in the context of air pollution controls, and 50% lower in the contexts of improved birth control pills and reduced medical radiation. However, nearly a quarter of respondents — 12 out of 52 — gave equal valuations to each safety improvement in each elicitation question. Six of these 12 subjects only came to the view that “a life is a life” after discussions (five after unmoderated pair discussions and one after an interview with the moderator); one subject rejected this view following discussion unmoderated.

Individual ratios were estimated as follows. Denote the $i$th individual’s relative valuation of public safety in area $X$ by $\lambda_i(X) = V_i(X) / V_i(C)$. The $V_i(\cdot)$ represent individual marginal rates of substitution of wealth for public safety. $^5$

The WTP estimate of $\lambda_i(X)$ is the $i$th individual’s WTP in area $X$ divided by her WTP in area $C$. The Relative Monetary Values estimate is her Relative Monetary Value in area $X$ divided by £100 m. The Matching estimate is her stated ratio, $\Delta C / \Delta X$, where $\Delta C$ represents the expected number of deaths prevented in area $C$, and $\Delta X$ the number in area $X$. Since the individual ratios were approximately lognormally distributed, the relative values were estimated using geometric means (see Appendix C).

The aggregate rankings of the six improvements (both means and medians) were precisely consistent across all four elicitation questions. At the individual level, however, most individuals exhibited some internal inconsistency between their rankings from the four different elicitation questions. Only 19 out of 51 respondents (37%) gave final answers which implied precisely consistent rankings (including ties) between all four elicitation questions, and 12 of these expressed equal valuations for each improvement in each question.

Table 2 shows the average willingness to pay figures. Most of the WTP responses were round numbers between £1 and £50. In view of this, it should be

$^5$ Individual marginal rates of substitution regarding private or consumer health risk, which has been the focus of most previous studies, may differ from those regarding public health risk, which is the focus of this study. This is because individual preferences may be influenced by (perceptions of) the health risks faced by other people.

noted that respondents were not presented with any monetary “cues”. All of the WTP questions were presented in tabular format on the same page, without the use of payment cards, and each response was written in two blank columns headed “£” and “p” (see Fig. 4).

Although the study was intended to elicit relative valuations, it is worth considering the “absolute” monetary values of life implied by the WTP responses. We can make a “quick and dirty” calculation of this kind if we assume that the sample is representative, and that household WTP each year to save 100 lives over 10 years translates into household WTP to save ten lives this year. Given that there are about 25 million households in the UK, monetary values of life for the UK population can be estimated from Table 2 by multiplying each WTP figure by 2.5 million. The resulting values of life lie between £30–110 m for arithmetic means and £12–25 m for medians. These values are an entire order of magnitude higher than most standard value of life estimates from road safety and occupational safety studies, which typically lie in the range £1–4 m (Beattie et al., 1998a). One possible explanation is that this study presented information in terms of lives saved rather than probabilities of dying. Another possible explanation is that this is due to a scope effect, as this study used a smaller magnitude of risk reduction than most previous studies. In this study, the risk reduction was 10 in 25 million per year, which means that a willingness to pay of £1.00 would be required to yield a value of life of £2.5 m.

6. A normative classification of respondents’ stated reasons for preference

The qualitative data from the group discussions, pair discussions and feedback interviews gave a clear indication of the reasons for subjects’ relative valuations. These reasons have been classified into three principle categories: (i) “psycho-social factors” relating to perceived characteristics of the health risk, (ii) “socio-economic factors” relating to observable characteristics of the affected population, and (iii) “confounding factors” that appear to represent confusions or misunderstandings about the purpose of the study.

This is an explicitly normative classification scheme, built up from the data with explicit reference to the theoretical purpose of this study: namely, to measure “true” economic preferences that represent individual well-being. This normative approach to analysing qualitative data differs from the “grounded” approach often used in psychological and sociological research, which classifies the data without explicit normative reference to pre-existing theory (Pidgeon and Henwood, 1992). Subjects’ reasons for their valuations have also been related to the well-known psychometric factors from risk perception research — “Dread”, “Unknown”, and “Number” — which have already been described in Section 3.

In principle, it would be possible to estimate the strength of preference associated with each factor. For example, individual valuations could be regressed.
against individual ratings of particular factors (as was done, for example, by McDaniels et al., 1992), or against binary variables that represent whether or not the factor was mentioned by that individual (as was done, for example, by Schkade and Payne, 1994). However, this approach might be too crude to cope with the complex intercorrelations between different factors in different contexts. And, in any case, “ratings” and “mentions” of particular factors are a rather fragile basis for quantitative work.

By far, the most frequently cited reason that subjects gave for valuing safety differently across the six areas was the degree of “choice” and “control” they felt themselves to have — i.e., the perceived extent to which they could (a) choose to be exposed to the hazard (for instance, by owning a car) and (b) control their own level of risk from exposure to the hazard (for instance, by driving carefully). Almost all of the subjects mentioned one or both of these factors as primary considerations during their pair discussion or feedback interview.

Subjects generally agreed that the degree of choice and control was lowest in area A (air pollution) and highest in areas B and M (birth control pills and medical radiation). However, subjects often disagreed with one another about the relative degree of choice and control involved in areas C, F and R (car accidents, food poisoning and rail accidents). A handful of subjects also mentioned a particularly high degree of worry about areas A and C, often in connection with the safety of their children. However, none of the respondents mentioned factors relating to the process of dying in different contexts, or said that they would rather die one way rather than another.

These three psycho-social factors — choice, control and worry — all appear to be intelligible reasons for placing a higher value on saving life in one area rather than another. Although these factors have not traditionally been incorporated into welfare economic models, they can all be understood from a welfare-theoretical point of view. Greater perceived freedom to choose and to control the risks that you face, and reduced worry, may improve your quality of life even if it actually reduces your life expectancy (as estimated by yourself or by the health experts). So you may be prepared to trade off the value of living longer against the value of living with greater freedom from bureaucratic controls and with greater peace of mind.

All three of these contextual factors are closely related to the first of the three primary psychometric factors from risk perception research, namely the “Dread” factor. Interestingly, none of the subjects appeared to be influenced by considerations relating to the second psychometric factor, “Unknown”. Of course, since

---

6 The factor that respondents referred to as “choice” is usually referred to in the risk perception literature as “voluntariness”.

the six policy areas were originally selected for this study as involving familiar risks, subjects may have perceived little difference between them along the “Unknown” dimension.

Only a handful of subjects mentioned any of the socio-economic factors that have traditionally been incorporated into economic models. One was what subjects termed the “selfish” reason of being personally less at risk from some health risks than others. If individuals are more concerned with their own health than with other people’s, we might expect that relative valuations would depend upon expected variations in the magnitude of private risk reduction for different individuals and households in different areas. A few subjects mentioned this as a reason for placing a lower value on rail safety than road safety, since they were infrequent rail users. This may explain why — unlike Jones-Lee and Loomes (1995) which selected subjects who were frequent rail users — this study found no quantitative evidence for a premium on rail safety as compared to road safety.

Another potentially relevant socio-economic factor was the difference in age and life expectancy of those most at risk. Economic models that set positive value on longer life expectancy would predict that health improvements that mainly benefit those with high life expectancy (e.g., young and healthy people) should receive a higher valuation than ones which mainly benefit those with low life expectancy (e.g., elderly and infirm people). Information about the age groups most at risk was presented to subjects on the information cards — in particular, it was stated that the elderly and infirm were particularly at risk from food poisoning and air pollution. It was surprising, therefore, that only one of the subjects spontaneously raised the issue of life expectancy or age in discussions although a handful of subjects said that they were particularly concerned when their own children were at risk.

Several subjects mentioned various apparently confounding factors that are hard to square with the stated purpose of the study. Some mentioned ongoing benefits beyond the stated time horizon of 10 years, and associated non-fatal detriment to health as well as the stated number of deaths (especially in connection with air pollution). A handful were more concerned with the costs of the improvements rather than the benefits. One or two did not believe that 100 deaths could be prevented over the next 10 years from either rail accidents or medical radiation. These factors are all incompatible with the stated purpose of the study in a quite straightforward sense.

Some subjects also mentioned factors which are incompatible with the purpose of the study in a rather less straightforward sense. For instance, in line with previous studies (see, especially, McDaniels et al., 1992), several respondents were concerned about the baseline probability of death from a particular hazard as opposed to the stated change in the probability of death. More specifically, subjects mentioned gave reasons relating to the overall importance of a particular area of policy. This directly related to the third primary psychometric factor from risk perception research, the perceived “Number” of people exposed to risk.
Now, to some extent, this consideration may feed into the anxiety that people feel. It may be worth spending money on high profile policy areas to help to allay people’s fears, even if that expenditure is not cost-effective. However, it seems likely that this factor also represents a degree of confusion in subjects minds between levels and changes of probability, due to their unfamiliarity with risk concepts and economic concepts. That is, it is hard to understand why a concern for the existing total number of lives lost should influence the individual’s valuation for saving one more (marginal) life. In other words, why should the composition of an individual’s probability of dying matter, so long as both her total probability of dying and her quality of life remain the same? So this factor may partly reflect a misplaced concern based upon a misunderstanding of the way in which expressed valuations would ultimately be used to guide policy.

Some of the subjects also mentioned political or ‘symbolic’ factors that would appear to have little or nothing to do with the health as such — for example, the desirability of protecting the environment and the desirability of a shift towards public transport and away from private transport. This is a natural concern for subjects to have in relation to public policy. However, it is hard to see how these considerations relate to the health component of an economic valuation. Rather, these considerations would normally enter the analysis at some other stage, as separate costs and benefits of the policy proposal in question. So, as before, these considerations appear to be misplaced once one appreciates the purpose of the study and the practical uses to which its results will be put.

7. The effects of elicitation question mode on responses

The elicitation question modes had two systematic effects on responses. First, questions framed in terms of lives saved (i.e., Matching questions) elicited a narrower range of cardinal relative valuations than those framed in terms of money (i.e., WTP and Relative Monetary Values questions). Second, Relative Monetary Values questions tended to elicit higher valuations relative to the baseline context of $C$ than WTP questions.

The first of these effects is particularly interesting and puzzling since it suggests that the currency of money may be perceived differently than the currency of lives saved, even though both currencies buy the same size of health improvement. In all three contexts valued using Matching questions (C, R and F) the distributions of individual relative values from the Matching questions were narrower than the corresponding distributions from the WTP and Relative Monetary Values questions. If we allow for an independently distributed degree of random noise or error in individual responses between different questions, we can test whether or not this observed difference in responses could be expected as the result of chance. This was tested using a Moses test of extreme reactions,
Fig. 6. Matching vs. WTP ratios (comparing car accidents with food poisoning).

conditional on placing different valuations on both groups. The effect was significant at 5% for the comparison between C and F, although not for the comparisons between F and R or F and C. Fig. 6 illustrates this pattern for car accidents vs. food poisoning, by comparing the distribution of individual relative values implied by Matching responses with the corresponding distribution implied by WTP responses.

This ‘currency effect’ can also be seen by computing the WTP amount implied by a Matching response in conjunction with stated WTP in one of these two contexts. Formally, the $i$th individual’s implied WTP for a particular improvement (say improvement X) is defined as $\text{WTP}_i(C) \times \lambda_i(X)$, where $\text{WTP}_i(C)$ is her stated WTP for improvement C and $\lambda_i(X)$ is the Matching estimate of her valuation of X relative to C. A ‘‘currency effect’’ would pull Matching responses

---

7 The group of respondents who took the view that ‘‘a life is a life’’, and gave equal valuations to all health improvements, form a distinctive modal response category and we would not expect them to be influenced in the same way as other respondents by the response mode.
in towards the value given to the numeraire of $C$, and push WTP responses away from that numeraire value. This means that implied WTP would be pulled downwards relative to stated WTP for improvements valued more highly than $C$ (like improvement $F$) and upwards for improvements valued less than $C$ (like improvement $R$). This is indeed what we observe: implied WTP tended to be higher than stated WTP in context $R$ and lower in context $F$. A Wilcoxon signed-rank test showed that this disparity was significant at 5% for context $R$ but not for context $F$. The scatter plot in Fig. 7 shows this effect, by comparing stated and implied WTP for the rail safety improvement.

A second effect of elicitation question framing was that the Relative Monetary Values questions tended to elicit higher valuations relative to the baseline context of $C$ than the WTP questions. More specifically, geometric mean relative valuations for improvements $B$, $M$ and $A$ were respectively 13, 6 and 47 percentage points higher under Relative Monetary Values questions than WTP questions (although $F$ and $R$ were generally valued about the same as $C$ in both questions). These differentials in valuations of $B$, $M$ and $A$ between the two questions were individually significant at the 5% level in Wilcoxon signed-rank tests, although this effect did not significantly alter the aggregate rankings of these improve-

---

8 However, Relative Monetary Values valuations for improvements $F$ and $R$ do come out as higher than the corresponding WTP valuations if arithmetic means are used as the measure of central tendency.
ments. One possible explanation for this effect is that there may have been a form of "symbolic bias": respondents may have stated Relative Monetary Values higher than the baseline figure of £100 m for improvement C in order to symbolise the view that more money should be spent on saving lives in general.

8. The effects of deliberation and discussion on responses

Deliberation and discussion had two main effects on responses: (i) an increase in the degree of consistency between individual answers to different elicitation questions, and (ii) higher valuations for health interventions A, F and R. This first effect can be examined by looking at the proportion of subjects giving precisely consistent individual rankings across the four elicitation modes (in terms of strict preference and indifference). This proportion rose from 17% to 37% following the 20-min pair discussions and feedback interviews. Setting aside the 12 respondents who gave equal valuations for all areas of safety in their final answers, this increase is from 2 out of 39 up to 8 out of 39, of whom 5 took part in an unmoderated pair discussion and 1 in an interview. These individual-level inconsistencies averaged out in aggregate, so that aggregate rankings for the group as a whole were precisely consistent across different questions.

Table 3 illustrates the general increase in improvements A, F and R following discussions, by showing the change in arithmetic mean WTP from initial to final responses. The significance of these changes can be examined using standard individual-level Wilcoxon signed rank tests, using a 5% level of significance.9 The increases in F and A are significant in terms of absolute WTP valuations, although the corresponding WTP rankings and relativities were not significantly affected. Turning to the Matching questions, there was a significant shift in relative value towards R and away from C.10 Finally, looking at the Priorities questions, intervention A moved significantly upwards and M downwards in the rankings (sign tests on difference between initial and final rankings). The Relative Monetary Values question was not administered in the initial set of questions.

The effects of variations in the "warm up" question and in the number of policy contexts being valued are also consistent with the hypothesis that health intervention R received a higher valuation following extra deliberation. The

9 These tests assume that the response error is independently distributed among individuals and groups. If response error were perfectly correlated within groups, then each group should instead be treated as a single data point. However, although there were some signs of "group dynamics", there was always considerable variation in responses between individuals in each group and insufficient correlation to justify treating the group as the unit of analysis.

10 This was due to a significant shift in responses to the preliminary pairwise choice question, rather than a shift in responses to the follow-up trade-off question.
relative value of improvement R was significantly higher in WTP and Relative Monetary Values questions for respondents who discussed this context in the preliminary warm up discussion, and in WTP questions for those who valued only four improvements, rather than all six (Kruskall–Wallis tests, using a $2 \times 2$ block design with nine observations in each block). Arguably, preliminary discussion of context R is a form of additional deliberation, and those who valued only four improvements had additional time to deliberate about each improvement than those who valued all six.

Finally, it is worth noting that there was some clustering of similar individual responses within particular groups (especially over the question of whether or not to give equal valuations to each health improvement). This suggests that the dynamics of group discussion may have had effects which differed from one group to another. However, there is evidence that this clustering of responses was not the result of a “dominant” group member dictating a joint response. If that was the case, one would expect individual responses to the Relative Monetary Values question made in secret to differ from the joint response to this question, made under the watchful eye of the dominant partner. But in virtually every case in which individuals gave the same individual response to this question (16 out of 17 individuals), both partners gave this same individual response.

9. Discussion

This section discusses three main issues: (i) the weight given to psycho-social considerations as opposed to health gains, (ii) the effect of framing elicitation questions in terms of lives saved rather than money, and (iii) the influence of deliberation and discussion.

The respondents in this study gave weight to psycho-social considerations relating to choice and control by those at risk, over and above the value of saving as many lives as possible. In response to four different elicitation question modes, both before and after pair discussions, they consistently valued saving life more highly in policy contexts involving a lower degree of choice and control. This

---

Table 3
Change in willingness to pay following discussion (arithmetic means)

<table>
<thead>
<tr>
<th></th>
<th>C (n = 51)</th>
<th>F (n = 51)</th>
<th>B (n = 27)</th>
<th>M (n = 27)</th>
<th>A (n = 51)</th>
<th>R (n = 51)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>Initial WTP</td>
<td>£22.48</td>
<td>£27.73</td>
<td>£11.19</td>
<td>£15.60</td>
<td>£27.93</td>
</tr>
<tr>
<td>(2)</td>
<td>Final WTP</td>
<td>£24.88</td>
<td>£31.81</td>
<td>£13.11</td>
<td>£14.05</td>
<td>£43.74</td>
</tr>
<tr>
<td>Change in WTP (2)−(1)</td>
<td>£2.40</td>
<td>£4.08</td>
<td>£1.92</td>
<td>£1.55</td>
<td>£15.81</td>
<td>£3.99</td>
</tr>
</tbody>
</table>
finding cannot be attributed to straightforward errors or misunderstandings on the part of respondents, given that respondents took part in a moderated 20-min warm-up discussion explaining the purpose of the questions and an unmoderated 20-min pair discussion about their initial answers to the questions.

Nor can this finding be attributed to differences in the life or health expectancy of those whose lives might be saved. The two contexts which respondents were told posed special risk to the elderly and the infirm — food poisoning and air pollution — in fact received among the highest relative valuations of life. And, more generally, age and life expectancy were rarely discussed: the issue was spontaneously mentioned only once, whereas psycho-social issues of choice and control were spontaneously raised in all 10 groups.

Furthermore, in response to the attitude question, “Public safety expenditures should be lower when those most at risk of dying are the elderly”, only 6 of the 52 respondents agreed and 25 strongly disagreed. This contradicts other survey findings which suggest that members of the public, young and old alike, feel that the young should be given priority in the allocation of health care resources (for example, Charny et al., 1989). One possible explanation is that this is due to question framing. An alternative framing, for instance, would ask whether “Public safety expenditures should be higher when those most at risk of dying are the young”. Another possible explanation is that people’s preferences regarding the allocation of resources between health care treatments differ fundamentally from their preferences in the realm of preventative health and safety measures.

Turning to the influence of question framing on responses, one of the most interesting and puzzling findings in this study was that elicitation questions framed in the currency of lives saved (i.e., Matching questions) elicited narrower range of relative values than elicitation questions framed in the currency of money (i.e., WTP and Relative Monetary Values questions). One possible explanation for this “currency effect” is in terms of the “anchoring and adjustment” heuristic (Kahneman and Tversky, 1982). In questions involving lives saved, respondents may take equal valuation of lives as the “anchor” from which adjustments are made for other considerations. When dealing with the currency of money, by contrast, spending more money on more important areas of safety may be the “anchor”, and adjustments made from this starting point.

Whatever its cause, this “currency effect” presents researchers with a potential dilemma, as it means that different elicitation questions may have different implications for the distribution of resources between life saving programmes. It might be argued that the currency of lives saved is more appropriate, on the grounds that it forcefully presents the opportunity costs of spending money in different areas. Furthermore, many of the respondents said that they found it easier to answer questions framed entirely in terms of lives saved, rather than to compare lives with money. On the other hand, it might be argued that the currency of money is more appropriate, since the currency of lives saved gives too much prominence to the idea that everyone’s life is equally valuable. The idea that
everyone should be treated equally is essentially a notion of equity, which arguably should not be given prominence in a valuation task designed to measure efficiency.

Finally, we turn to the effects of discussion and deliberation. We can expect some changes in responses due to random noise. In standard pairwise choice laboratory economic experiments, for instance, about 25–30% of subjects reverse their responses when asked the same question twice (Camerer, 1995). Interestingly, a similar proportion of subjects (one-third) changed their responses in this experiment. However, these changes were not simply random noise. There were two systematic effects: first, an increase in consistency between responses under different elicitation modes, and second, an increase in valuations of health interventions A, F and R.

One possible explanation of this latter effect is that this is simply a “herd effect” of moderation of responses towards the mean response, due to social pressures to conform. An alternative explanation is that deliberation and discussion led respondents to gain an increased awareness of the contextual factors of choice and control, which are reasons stated by respondents for giving A, C, F and R higher valuations than M and B. On this interpretation, it might be argued that discussion and deliberation had some success in “improving” individual preferences, and moving them closer to “true” economic preferences.

In a similar vein, the increase in internal consistency following discussions might also be interpreted as an “improvement” in preferences. However, individual responses continued to exhibit a substantial number of inconsistencies between different elicitation questions. Furthermore, the qualitative evidence suggested that a minority of subjects persisted in taking into account apparently confounding factors which appear to reflect misunderstandings about the nature and purpose of the elicitation questions.

This was despite the 20-min warm up discussion, and strenuous efforts to design the questionnaire so as to minimise misunderstandings. In relation to the political factors, for example, subjects were instructed to, “please focus on improvements in safety rather than other kinds of improvement, such as reduced traffic congestion”, and in each question respondents were asked to assume that the safety improvement “would have no other significant effects” [bold in original]. During feedback interviews and in written comments, however, some respondents said that they tried to set aside these confounding factors but felt that they were unable to do so.

One possible way to reduce inconsistencies and misunderstandings still further would be to engage subjects in a more formal procedure of decision analysis, over a period of days or weeks rather than a period of hours (as recommended, for example, by Gregory et al., 1993). However, the clustering of individual responses by pairs and by groups should sound a note of caution. Research into group psychology suggests that highly intensive group discussions may produce final answers that are more consistent and more confidently expressed than initial
answers — and yet no closer to the truth (see, for example, Irwin and Davis, 1995). More generally, a lengthy and formal procedure of decision analysis increases the danger that subjects will be brow-beaten or indoctrinated in one way or another.

10. Conclusion

After careful consideration in group discussions, the respondents in this study were prepared to sacrifice statistical lives for the sake of psycho-social benefits associated with the degree of choice and control that people have over different risks to health. Such concerns were sufficiently strong for these respondents to value life saving differently across six different policy contexts by a factor of ± 50%. Contrary to findings in other studies, this study found that the life expectancy and age of those at risk were not important factors in determining individual preferences between health interventions.

From a methodological point of view, perhaps the most important finding was that elicitation questions framed in the currency of lives saved (“Matching” questions) elicited smaller variations in the value of life than those framed in the currency of household income and public spending (“Willingness to Pay” and “Relative Monetary Values” questions). This interesting and puzzling “currency effect” poses a potential dilemma, since the different frames have different implications for allocation of resources between different life saving programmes. It was also found that the Relative Monetary Values elicitation question tended to elicit higher values (relative to the baseline context) than Willingness to Pay questions.

A second key methodological finding was that 20-min pair discussions reduced (but did not eliminate) internal inconsistency between responses to different elicitation questions. Furthermore, these pair discussions also increased the valuation given to contexts involving a low degree of perceived choice and control, suggesting that discussions may have helped to highlight the importance of these psycho-social factors.

In conclusion, this study finds evidence that (i) the influence of psycho-social factors on individual preferences for health interventions remains strong (and may even increase slightly) following in-depth deliberation and group discussion, and (ii) the influence of psycho-social factors on individual preferences is less strong when elicitation questions are framed directly in terms of lives saved rather than money.

Acknowledgements

The research was supported by the ESRC, under postgraduate research studentship award R00429424071 and research award R000234987 “Theoretical and Empirical Issues in the Valuation of Safety”. The author is particularly grateful to
Graham Loomes for advice in designing this study, and to John Bone, John Hey, Robert Sugden and two anonymous referees for detailed and extremely helpful comments. Thanks also to Judith Covey and Anne Spencer for help with the survey design, and Melinda Acutt, Gabriella Berloffa, Diane Dawson, Paul Dolan, Richard Fordham, Martin Forster, Mike Jones-Lee, Pamela Mason, and Alan Williams for useful comments. Errors and opinions are the author’s own responsibility.

Appendix A. Information cards describing the six life saving contexts

A.1. Improvement R

100 fewer deaths over the next 10 years from rail accidents
- Risk of dying stems from both train crashes and underground fires
- All rail and underground passengers and workers are at risk of dying
- This category does not include rail suicides, children playing on the line, accidents while trespassing, or accidents at level crossings

A.2. Improvement C

100 fewer deaths over the next 10 years from car accidents
- All car or van drivers, and their passengers, are at risk of dying
- This category does not include pedestrians, cyclists, or people on buses, coaches and lorries

A.3. Improvement A

100 fewer deaths over the next ten years from air pollution
- Risk of dying stems from both lung diseases and heart diseases
- Those most at risk of dying are people with existing lung or heart conditions, especially the elderly
- Most all pollution in Britain is caused by road traffic
- Masks may have little protective effect

A.4. Improvement B

100 fewer deaths over the next ten years from birth control pills
- Risk of dying stems from both blood diseases (thrombosis, strokes and heart attacks) and cancers
• All types of birth control pill increase the risk of dying

• Pill-takers most at risk of dying are smokers, overweight people, and people with a family history of thrombosis or cancer

A.5. Improvement F

100 fewer deaths over the next ten years from food poisoning

• Risk of dying stems from various poisons and bacterial infections in food

• Those most at risk of dying are people with a weak immune system, especially the elderly, the very young, and those with chronic illnesses

• Careful food handling reduces the risk of dying but cannot eliminate it

• This category does not include BSE ("Mad Cow Disease"), cancer-inducing pesticides or pollutants, or dangers from genetically modified organisms

A.6. Improvement M

100 fewer deaths over the next ten years from medical radiation

• Risk of dying stems from an increased likelihood of cancer later in life

• All those exposed to medical radiation from medical diagnoses or treatments, including routine X-rays, are at risk of dying

• Those most at risk are people exposed to lengthy and repeated doses of medical radiation, especially those undergoing radiological treatment for major illnesses

• This category does not include natural radiation, nuclear power, or nuclear weapons

Appendix B. Questionnaire preamble shown to all respondents

B.1. Introduction

B.1.1. Aims and objectives

• This workshop aims to find out your opinions about the relative value of public safety improvements in different areas of safety

• It involves both written answers to various questions and tape-recorded discussions
B.1.2. Agenda
- Group Discussion (20 min)
- Initial Questions (35 min)
- Break for Light Refreshments
- Discussions in Pairs (25 min)
- Repeat Questions (20 min)

B.1.3. Instructions
- Please write your answers in the shaded boxes
- Please complete the questions on your own
- Please ask whenever any question is at all unclear
- Please wait until everyone has finished before turning to the next question
- Please do not turn back to previous questions

B.1.4. Things to bear in mind
- The questions ask you to imagine several different public safety improvements, paid for by the UK taxpayer, which are all based on *real possibilities*
- Most of these improvements would benefit almost everyone in the UK, *including yourself*, in the sense that the risk of dying in the next 10 years would be reduced
- You are asked to assume that each improvement would *definitely* prevent a total of 100 deaths in the UK over the next 10 years
- When answering the questions, please focus on improvements in *safety* rather than other kinds of improvement, such as reduced traffic congestion
- Please remember that there are *no right answers* to any of the questions

Appendix C. Which statistical measure of central tendency for relative values?

The estimation of (aggregate) relative values raises a statistical difficulty, since it is not clear which measure of central tendency is appropriate for ratios. I have chosen to use *geometric* means. However, this needs some justification, since Jones-Lee and Loomes (1995) have argued in favour of estimating relative values using *arithmetic* means.
According to standard economic theory, the monetary value of statistical life is equal to the population arithmetic mean marginal rate of substitution (MRS) of wealth for risk of dying for the affected individuals. Let \( y_i \) denote the \( i \)th individual’s MRS of wealth for health risk in policy context \( y \), and \( x_i \) her corresponding MRS in context \( x \). We can write the appropriate monetary values of life by \( E(y_i) \) and \( E(x_i) \), respectively, where \( E(.) \) denotes the expectations operator. The relative value of life in context \( y \) with respect to context \( x \) is then \( E(y_i)/E(x_i) \). The problem is to determine which measure of central tendency should be used for a sample of individual relative valuations, \( y_i/x_i \), in order to estimate this aggregate relative value.

Jones-Lee and Loomes (1995) address this problem by assuming that the underlying distribution of preferences can be represented by the following simple model:

\[
y_i = k x_i u_i
\]  

(1)

where \( k \) is a positive constant and \( u_i \) is a random term, independent of \( x_i \), which represents variation in preferences between individuals. Under the further assumption that \( E(u_i) = 1 \), they show that both the population arithmetic mean relative valuation, \( E(y_i/x_i) \), and the ratio of population arithmetic means, \( E(y_i)/E(x_i) \), are exactly equal to \( k \). Under these conditions, the arithmetic mean of \( y_i/x_i \) is then clearly the appropriate measure of central tendency.

However, these conditions do not appear to hold for the data in this study, which have two salient characteristics. First, the ratios appear to be roughly lognormally distributed. Second, for each pair of contexts, the arithmetic mean of \( y_i/x_i \) is consistently higher than the inverse of the arithmetic mean of \( x_i/y_i \).

Taken together, these two characteristics are incompatible with the assumption that \( E(u_i) = 1 \).

From Eq. (1) we can write

\[
E\left( \frac{x_i}{y_i} \right) = E\left( \frac{1}{k u_i} \right)
\]

(2)

\[
E\left( \frac{x_i}{y_i} \right) = E\left( \frac{1}{u_i} \right) \frac{1}{k}
\]

(3)

But if \( u_i \) is lognormally distributed then \( E(1/u_i) = E(u_i) = 1 \), and hence

\[
E\left( \frac{x_i}{y_i} \right) = \frac{1}{k}
\]

(4)

But then, \( E(y_i/x_i) = k \) should be exactly equal to the inverse of \( E(x_i/y_i) \), not consistently higher.

Given the original model, an assumption that is compatible with the two characteristics is that \( E(\ln u_i) = 0 \). This new assumption implies that \( E(u_i) = \)
\[ E(1/u_c) = \exp(\sigma^2/2) \] will be greater than one, where \( \sigma^2 \) is the (positive) error variance. Now, from Eq. (1)

\[ E\left( \frac{y_i}{x_i} \right) = k \, e^{\frac{\sigma^2}{2}} \]  

(5)

and from Eq. (3)

\[ \frac{1}{E\left( \frac{x_i}{y_i} \right)} = k \, e^{\frac{\sigma^2}{2}} \]  

(6)

Clearly the value of the expression in Eq. (5) will always be greater than \( k \), and that of the expression in Eq. (6) will be always be less than \( k \), in accordance with the second characteristic of the data. But, if this new assumption is true, the estimated relative valuation will either be greater than one or less than one, depending which way around the arithmetic mean is computed. So is not clear whether to estimate \( E(y_i)/E(x_i) \), using the sample arithmetic mean of \( y_i/x_i \), or to estimate \( E(x_i)/E(y_i) \), using the sample arithmetic mean of \( x_i/y_i \). The theoretically correct measure depends upon an entirely arbitrary decision about which way around the two risks of dying are modelled, that is, whether the model is specified as \( y_i = kx_i, \mu_i \) or \( x_i = ky_i, \mu_i \).

For data with these two characteristics, therefore, the geometric mean — which is equal to \( k \) and which lies between the two possible measures using arithmetic means — would seem to be a more appropriate measure of central tendency. In principle, this conclusion may sometimes apply to the estimation of “absolute” monetary values as well as relative values, since a monetary value is also a trade off ratio between two goods — namely, wealth and health risk. But in general, of course, since the underlying distribution is unknown, the appropriate measure must be a matter for case by case judgement, taking into account the general characteristics of the data.

References

Beattie, J., Covey, J., Dolan, P., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N., Robinson, A.,


