*Applied Economics*, 1995, **27**, 1103–1111

# *Risk–risk versus standard gamble procedures for measuring health state utilities*

PAUL DOLAN, MICHAEL JONES-LEE and GRAHAM LOOMES‡

*Department of Economics, University of Newcastle, Newcastle NE1 7RU, UK* and ‡*Department of Economics, University of York, Heslington, York YO1 5DD, UK*

Contingent valuation (CV) has been widely used to measure the potential benefits derived from different policy decisions. However, doubt now exists about the validity of the CV method and alternative approaches to benefit valuation have been proposed. The paper reports on the results of a study which was designed to test the viability of two of the most prominent of the alternatives: the risk–risk (RR) and standard gamble (SG) approaches. If individual preferences are consistent with the axioms of von Neumann–Morgenstern expected utility theory (EUT) then the two methods should generate the same interval scales for any given set of health states. However, the results show that SG utilities are substantially higher than RR ones, thus casting doubt on these axioms. The paper discusses alternatives to EUT which might better explain the discrepancies found. It also considers whether the results might be explained in terms of status-quo bias and/or by the relative difficulty of RR questions. The results presented may have important implications for other areas of applied research in which there exists uncertainty about outcomes.

## I. INTRODUCTION

For several years it was widely believed that contingent valuation (CV) questions offered the most direct and effective means of establishing preference-based monetary values of health and safety. For example, the willingness-to-pay (WTP) value for the prevention of a road fatality adopted by the UK Department of Transport (DoT) in 1988 was based in part on the findings of a CV study (see Jones-Lee, 1989a).[1] However, the CV method has recently become the subject of a heated controversy[2] and there are now serious doubts as to the reliability of monetary values of health and safety established by this method

In view of all this, it is not surprising that many researchers have sought alternative methods of measuring preferences over health and safety. Two of the more prominent and apparently effective of these methods are the risk–risk (RR) and standard gamble (SG) approaches.[3] Therefore, when the DoT commissioned a programme of research to put the valuation of preventing non-fatal road injuries on the same preference-based methodological footing as that employed in the valuation of preventing road fatalities, it was decided to explore the use of RR and SG methods as well as direct CV questions

The main purpose of the study reported in this paper was to test the viability of the RR and SG procedures, i.e.

[1]The WTP figure adopted was £500 000 in 1987 prices, compared with a figure of about £190 000 based on the earlier gross output approach For a discussion of the principles underpinning WTP and alternative procedures for valuing safety, see Jones-Lee (1989b)
[2]In the environmental context, possibly the most wide-ranging and fundamental debate, involving some fierce exchanges of views, followed the *Exxon Valdez* oil spillage: see Carson *et al* (1992), Arrow *et al.* (1993) and Hausman (1993).
[3]The RR method has been applied to the valuation of life and safety in the USA (see Viscusi, 1994) while the SG approach has been most widely used by health economists (see Gafni, 1994)

0003–6846 

to assess their acceptability to both interviewers and respondents, and to assess their effectiveness as a means of eliciting preferences. The strategy was to elicit utility indices for various serious injuries, taking the utilities of normal health and death to be 1 and 0 respectively. These utilities can then be used as the basis for estimating ratios of marginal rates of substitution of wealth for risk of non-fatal injury and wealth for risk of death.[4]

## II. THE RELATIONSHIP BETWEEN RR AND SG QUESTIONS

Both the RR and SG methods are built upon the same expected utility theory (EUT) foundations and thus *in principle* there is nothing to choose between using RR or SG questions, i.e. for any respondent, the two methods should produce the same utility for any given health state. This relationship between the methods can be neatly illustrated with the aid of a Marschak–Machina (MM) triangle which provides a way of representing the set of all certain and risky alternatives involving up to three mutually exclusive final outcomes (see Machina, 1987). If outcomes $x1$, $x2$ and $x3$ (where $x1$ is the least preferred and $x3$ is the most preferred) have probabilities $p1, p2$ and $p3$ respectively (where $p1 + p2 + p3 = 1$), then it is possible to describe any alternative as a pair $[p1, p3]$, and to represent that alternative as a point in a triangle like the one in Figure 1 where $p3$ is calibrated on the vertical edge and $p1$ on the horizontal edge. In the context of this paper, let the vertical edge of the MM triangle represent the probability of the most preferred outcome, normal health (denoted by J), and the horizontal edge represent the probability of the least preferred outcome, death (denoted by K). The certainty of the intermediate outcome, I, is represented by the point at the right angle of the triangle (i.e. where $pJ$ and $pK$ both equal zero).

For an expected utility maximizer, it is well established that preferences over different alternatives involving the three outcomes can be represented by a set of upward sloping parallel straight line indifference loci.[5] It follows that for an expected utility maximizer we need only identify two points that lie on the same indifference locus in order to determine the whole indifference locus (a straight line through those two points) and thus the full set of indifference loci within the triangle (all linear and parallel to the one initially identified).

Thus, so long as the axioms of EUT are satisfied, the only difference between RR and SG is that the two procedures operate in different areas of the triangle. To see this, consider Figure 2, which illustrates how the utility of injury state R might be elicited both by RR and by SG. The left-hand side of Figure 2 shows the top left-hand corner of the MM triangle magnified to allow us to see where the RR question is located. In the RR question relating to injury R, respondents were asked to consider a starting position where the current risk of death ($K$) is 80 in a million and the risk of injury $R$ is 120 in a million, so that the chances of being neither killed nor injured in the way described on card $R$ are 999 800 in a million. Thus the respondent is initially located at point A in Figure 2. She is then asked what decrease in the risk of death would be required in order to exactly compensate an increase of 60 in a million in the risk of $R$. Suppose the answer is that a reduction of 40 in a million in the risk of $K$ would be required. This would take the individual to point B, which means that under the assumptions of EUT, her indifference locus would be a straight line through points A and B, and the utility $U(R)$ would be computed as $(60 - 40)/60 = 1/3$ (more detail about the derivation of this value is given in the RR questions section below).

Now consider the same individual presented with an SG question. She is asked to imagine that she has suffered injuries which, treated in a particular way, will with certainty result in the health state described on card $R$. Thus she is initially located at point C in Figure 2. However, it is suggested that there is an alternative treatment which, if successful, will restore her to normal health, but if unsuccessful, will result in death. If she behaves according to EUT, she will be indifferent between the two alternative treatments when the second one offers a one-third chance of success and a two-thirds chance of failure – a prospect depicted by point D in Figure 2. The lines AB and CD are parallel and $U(R)$ again equals 1/3.

However, there are at least two important reasons why in practice the two methods may not produce identical results. First, there now exists a very considerable amount of experimental evidence (see Camerer, 1995) to suggest that the restrictions on behaviour imposed by EUT may be too severe for many people, and that even if indifference loci in the triangle are approximately linear,[6] they may be very far from parallel. What is less clear is the precise pattern and

[4]Of course, it is not always necessary to convert these utility weights into monetary values in cost-effectiveness and cost-utility analyses, it is sufficient to use the utility weights themselves to quantify all health outcomes In that some policy makers may be hesitant to base decisions on benefits denominated in monetary units, the SG and RR methods can avoid the political sensitivities of placing monetary values on health outcomes

[5]EUT incorporates the principles of betweenness and independence Betweenness entails that an individual who is indifferent between $X$ and $Y$ will also be indifferent between those alternatives and every probability mixture $\pi X + (1 - \pi) Y$, $0 < \pi < 1$ This ensures that indifference curves are linear Independence entails that if there is some third prospect, $Z$, then indifference between $X$ and $Y$ implies indifference between $\pi X + (1 - \pi)Z$ and $\pi Y + (1 - \pi)Z$ for all $\pi$ This ensures that all indifference curves in the triangle are parallel

[6]A number of experimental studies suggest that they are not linear – but there is no consensus about the nature of any non-linearities.

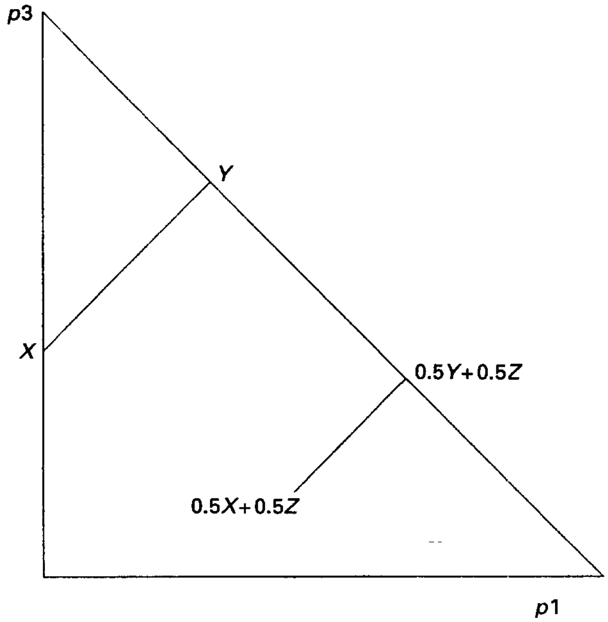


Fig. 1. *The above triangle illustrates the case where* $X = (x3, 0.4; x2, 0.6)$, $Y = (x3, 0.7; x1, 0.3)$, $Z = (x1, 1.0)$

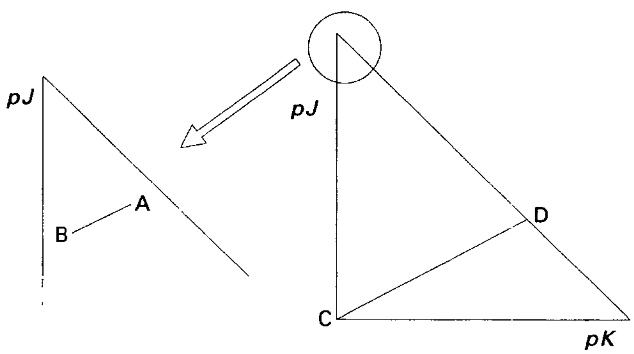


Fig. 2

extent of the failure of EUT in different areas of the triangle. For example, early experiments involving money payoffs focused on the bottom right-hand area of the triangle, and produced results which seemed consistent with indifference curves that 'fanned out' as if from some point to the south-west of the triangle.[7] However, as Camerer (1993) describes, various subsequent experiments have raised the possibility of more complex patterns, e.g. fanning out in the bottom right-hand corner but fanning *in* in the top left-hand corner, when all three payoffs were non-negative (the domain of gains), some 'reflection' of this pattern – fanning out top left, fanning in bottom right – when payoffs were non-positive (losses) and so on.

Second, one method may be easier for both respondents and interviewers to deal with. *A priori*, it was expected that the SG questions would be the more 'user friendly', for two reasons: they used probabilities expressed as chances out of 100, whereas the RR questions used the more unfamiliar and, arguably, less imaginable 'chances in a million'; and SG questions asked people to set a simple lottery equivalent to a given certainty, rather than try to match two risk changes moving in opposite directions. On the other hand, the RR questions involved probabilities, and variations thereof, that were much closer to the actual road risks that people face *ex ante*. Since the ultimate objective of the SG and RR questions was to provide the basis for estimating ratios of marginal rates of substitution (MRS) of wealth for risk of non-fatal injury and wealth for risk of death, with these MRS evaluated at existing levels of risk, then strictly speaking one should be operating in the upper left-hand corner of the MM triangle since, for the vast majority of people, road risks are very small in absolute terms. Thus indifference loci in the middle of the MM triangle, as elicited by SG questions, are useful only in so far as they provide accurate proxies for the loci in the top left-hand corner.

Therefore, it was also important that the pilot study assessed whether the two procedures tended to give similar results or whether they produced substantial (and systematic) disparities. If they do *not* produce sufficiently similar results, then it is necessary to ask whether the observed patterns are indicative of any particular non-EU models, or whether they might be explained by other factors, such as confusion.

## III. STUDY DESIGN

The study reported here was part of a larger design involving four different questionnaires. Two of these were based on variants of the standard CV questions, one was based on the RR method, and one on the SG method. The majority of interviews for all questionnaires were conducted by professional interviewers from National Opinion Polls Ltd (NOP), with 12 interviewers each carrying out at least 5 interviews of two types, to give an NOP quota sample of at least 30 for each questionnaire. In addition, three professional interviewers from the Transport Research Laboratory (TRL) each conducted two further interviews of each type, with the specific task of interviewing 'difficult' respondents (e g. those who were elderly and/or somewhat infirm). Thus, there were at least 36 interviews for each questionnaire.

The DoT defines a serious injury as one for which 'a person is detained in hospital as an in-patient, or any of the

[7]This evidence stimulated models such as Machina's (1982) generalized expected utility theory (GEUT) and Chew and MacCrimmon's (1979) weighted utility theory (WEU), which postulated that if fanning out occurred then it would do so across the whole of the triangle

following injuries whether or not he is detained in hospital: fractures, concussion, internal injuries, crushings, severe cuts and lacerations, severe general shock requiring medical treatment, injuries causing death 30 or more days after the accident' (Road Accidents Great Britain, 1989).[8] Clearly, this definition covers a very broad range of possible injuries, so prior to the construction of any questionnaires, it was necessary to devise a set of 'typical' serious injuries which spanned the range of severity. Seven such descriptions were printed on 15 cm × 10 cm 'shuffle' cards, to which a further three cards – representing normal health, death and 'whiplash' injuries[9] – were added. The set of ten cards, together with their code letters, are reproduced in Appendix A.

Respondents to all questionnaires were initially asked to rank the ten health state descriptions and then locate them on a hundred-point visual analogue scale (VAS) in such a way that the relative sizes of the intervals between locations indicated the relative extents to which each injury was regarded as worse than others. The purpose of this exercise was not principally to establish utility weights but to encourage respondents to think carefully about the health state descriptions and, given the relatively small size of the quota samples being interviewed, to provide a check on the comparability of the different subsamples. However, because it was felt that it would be too demanding and time consuming to ask RR or SG questions about all ten injuries, it was decided to focus on eliciting utilities for four of these ($W, X, S$ and $R$) and to use the information from the VAS exercise to indicate where the other injuries lay in relation to these, and to interpolate accordingly. The middle section of each questionnaire involved the RR/SG questions (see below) and a final section collected respondent background information

*Risk–risk (RR) questions*

This is a form of equivalence technique in which the respondent is asked the following kind of question: 'A policy will result in your risk of death being reduced by $x$ in a million but only by increasing your risk of a particular injury by $y$ in a million. Would you be willing to accept the policy?' (see Viscusi *et al.*, 1991) One of the numbers $x$ or $y$ can then be varied until the respondent is faced with a probability at which she is indifferent between the policy being adopted or not. In principle, by asking a series of such questions, all conditions can be related to each other on a utility scale.

With the utility indices for normal health and death respectively scaled at 1 and 0, it can be shown that for an expected utility maximizer the utility of a state $I$ is given by:

$$U(I) = (y - x)/y$$

where $y$ is the increase in the risk of injury $I$ and $x$ is the decrease in the risk of death that is needed to compensate for the increase (see, for example Jones-Lee, 1989a, 1989b; or Viscusi *et al.*, 1991).

In the RR questions, the increases in the risks of the various injury states were predetermined, and in each case the decrease in the risk of death was varied until indifference was reached. The increase in risk of injury state $R$ was 60 in a million, of $S$ was 120 in a million, of $X$ was 180 in a million, and of $W$ was 100 in a million (from initial risk levels of 120, 240, 360 and 200 in a million respectively). The decrease in the risk of death (from an initial level of 80 in a million) needed to compensate for these increases could take a number of values between 1 and 70 in a million. Appendix B shows an example of the type of question the RR respondents were asked.

*Standard gamble (SG) questions*

This method asks respondents to consider the choice between a certain prospect or an alternative with two possible outcomes, one of which (occurring with probability $p$) is better than the certain outcome and the other of which is worse. The aim of each SG question is to find the value of $p$ at which the respondent is indifferent between the certain prospect and the risky alternative

For an expected utility maximizer, the utility of a state I is given by:

$$U(I) = pU(J) + (1 - p)U(K)$$

where $U(J)$ and $U(K)$ indicate the utilities of states $J$ and $K$. When the best outcome is normal health, $U(J)$ is set equal to one, and when death is the worst outcome, $U(K)$ is set equal to zero, so that $U(I) = p$.

The sequence of SG questions was determined for each respondent by their ranking over the six health states $J, W, X, S, R$ and $K$. The 'better' lottery outcome was the highest ranked state, the 'worse' lottery outcome was the lowest (6th) ranked state and the certain outcomes were the states ranked 5th, 4th, 3rd and 2nd respectively. Appendix B shows an example of the type of question the SG respondents were asked

## IV. RESULTS AND DISCUSSION

Table 1 shows that the mean and median visual analogue scale scores of the key health states $W, X, S$ and $R$ of those

[8]In this paper, we shall not be concerned with 'slight' injuries questions about slight injuries were asked only in the two CV questionnaires
[9]These descriptions were prepared after consultation with Professor Charles Galasko and his team at the University of Manchester, who were undertaking research concerning the nature and frequencies of the full spectrum of road accident injuries

Table 1. *VAS scores for both subsamples*

| | Risk–risk ($n = 36$) | | Standard gamble ($n = 36$) | |
|---|---|---|---|---|
| Injury | Mean (SD) | Median (IQR) | Mean (SD) | Median (IQR) |
| *R* | 0.27 (0.27) | 0.25 (0.15–0.45) | 0.26 (0.28) | 0.25 (0.13–0.45) |
| *S* | 0.47 (0.26) | 0.47 (0.30–0.60) | 0.42 (0.22) | 0.45 (0.28–0.60) |
| *X* | 0.62 (0.20) | 0.65 (0.50–0.75) | 0.61 (0.18) | 0.67 (0.50–0.78) |
| *W* | 0.78 (0.14) | 0.80 (0.70–0.87) | 0.80 (0.15) | 0.85 (0.70–0.89) |

Table 2. *Utility indices for both subsamples*

| | Risk–risk ($n = 34$[a]) | | Standard gamble ($n = 27$[a]) | |
|---|---|---|---|---|
| Injury | Mean (SD) | Median (IQR) | Mean (SD) | Median (IQR) |
| *R* | 0.49 (0.40) | 0.58 (0.17–0.87) | 0.66 (0.28) | 0.73 (0.40–0.90) |
| *S* | 0.78 (0.21) | 0.88 (0.58–0.95) | 0.87 (0.15) | 0.93 (0.77–1.00) |
| *X* | 0.88 (0.13) | 0.94 (0.78–0.97) | 0.93 (0.12) | 0.99 (0.88–1.00) |
| *W* | 0.78 (0.24) | 0.90 (0.70–0.95) | 0.95 (0.12) | 1.00 (1.00–1.00) |

[a]Sample sizes vary from those in Table 1 because only cases where all VAS and RR/SG questions were answered are analysed.

who answered RR questions are very similar to the mean and median scores of those who answered SG questions. This suggests that in this respect the two subsamples are highly comparable. Table 2 shows the mean and median utility indices for these states as derived from the RR and SG methods.[10] It is immediately apparent that the utilities estimated via SG are higher for every level of injury than those elicited through RR.

If responses to both types of question accurately reflect respondents' true preferences, the evidence in Table 2 suggests that indifference loci through the middle of the MM triangle are consistently steeper than those in the top left-hand corner, as if, in this region of the triangle, indifference loci are fanning in. Such non-parallel indifference curves might be taken as evidence of non-EU behaviour, and would be consistent with models such as prospect theory (Kahneman and Tversky, 1979) or disappointment aversion (Gul, 1991) which suggest fanning-in in the top half of the triangle. If this were the case, the appropriate conclusion would be that relative to the indifference loci in the top left-hand corner of the MM triangle, SG responses overstate the utility indices for the various non-fatal injuries, or, put another way, SG responses understate the loss of utility from serious injuries which reduce a person's health state from normal health to anything from W to R and hence underestimate the ratios of marginal rates of substitution of wealth for the risk of these injuries relative to the marginal rate of substitution of wealth for risk of death.

This would seem to suggest one of two possible courses of action. If RR questions are at least as acceptable to interviewers and respondents as SG questions, then the former rather than the latter should be used. Alternatively, if RR questions are for various practical reasons simply much less manageable than SG questions, then the responses to SG questions may require some appropriate adjustment.[11]

However, there are at least two other possible explanations for the observed patterns of responses; namely, status quo effects and simple confusion.

### *Status quo effects*

The notion of a status quo effect is, essentially, that people may give some special emphasis to their current position, and react asymmetrically to movements away from that position, placing greater weight on what they perceive as losses *vis-à-vis* the status quo than on what they perceive as

[10]The majority of respondents to the SG questionnaire had a 'standard' ranking of states such that $J > W > X > S > R > K$ which means that for these respondents the tasks were presented in the same order as in the RR questionnaire. In the comparison between RR and SG, analysis is confined to those in the SG sample who had this standard ranking. Unfortunately, 5 such respondents were interviewed by an interviewer who failed to administer the SG questions correctly, so that these 5 also had to be excluded from the analysis, leaving a total of 27 usable sets of responses.

[11]Assuming linear indifference curves which all converge to some point to the north-east of the triangle, it is possible to calculate the maximum difference between the gradient of indifference curves in the top left hand corner of the triangle and the gradient of the curve through the origin. See Jones-Lee *et al.* (1993a, Appendix) for details.

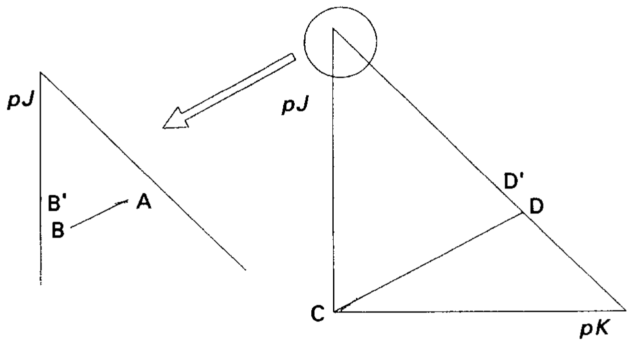


Fig 3.

gains. The frequently observed substantial disparities between what people say they would be willing to pay (WTP) for some marginal benefit, and what they would be willing to accept (WTA) as monetary compensation for a comparable marginal disbenefit, are often taken as evidence of such an effect (see Kahneman *et al.*, 1990; Viscusi *et al.*, 1987). In addition, the WTP–WTA disparity, and some evidence of the possible power of status quo effects in the context of road safety, have recently been reported by McDaniels (1992) and by Dubourg *et al.* (1994). How might such effects operate on the responses to RR and SG questions?

Consider Figure 3. Suppose that there were no status quo effects, so that a respondent's preferences conformed with EUT, as shown by the solid lines (which correspond to those in Figure 2). Now suppose that status quo effects are allowed to operate. In the RR question, the status quo is point A, and respondents are asked to say what gain (in terms of a reduced risk of $K$) will exactly offset a given loss (in terms of an increased risk of $R$). If perceived losses are weighed more heavily than perceived gains, the effect will be to elicit a larger requested gain than would otherwise be the case. that is, the response will suggest indifference between A and a point B′, where B′ lies to the north-west of B. In other words, the impact of such an effect in the context of RR questions will be to suggest indifference loci (such as the broken line from A to B′) that are flatter than would be the case if there were no effect. By contrast, the corresponding SG question 'endows' respondents with the certainty of the intermediate health state, i.e. starts them at point C, and then asks them to state the probabilities of gain (return to normal health) and loss (death) that would make them indifferent between C and a risky alternative. In this case, loss aversion would produce a response suggesting a steeper indifference locus than CD, e.g as represented by the broken line from C to D′ in Figure 3.

Thus, an alternative explanation of the disparities between RR and SG responses in Table 2, and the pattern of fanning in that they seem to entail, might be that the RR questions, as phrased in this study, asked people to consider a movement from north-east to south-west, whereas the SG questions involved a movement from south-west to north-east.

*Confusion and/or bias*

The discussion so far has been premised on the proposition that, in aggregate, respondents' answers reflected their true preferences more or less accurately. However, there appear to be some good reasons to doubt whether this was actually the case. For example, consider the mean VAS and RR values for injuries $X$ and $W$ (see Tables 1 and 2). By any standards, $X$ appears to be a worse injury than $W$ – it involves one to four weeks in hospital rather than two to seven days, and requires one to three years to return to normal health as compared with three to four months – and in the VAS exercise the mean (normalized) score for $X$ is 16 points lower than for $W$. But in Table 2, the ordering implied by the RR responses is reversed; mean $U(X)$ is 10 points *higher* than mean $U(W)$.

This discrepancy is not due to a small number of eccentric outliers. Analysis of the 34 individuals who answered all the relevant VAS and RR questions reveals that in the VAS exercise, 30 of the 34 respondents scored $W$ strictly better than $X$; but in the RR exercise, 11 said they would want a strictly greater reduction in the risk of death to compensate for a 100 in a million increase in the risk of $W$ than to compensate for a 180 in a million increase in the risk of $X$, and a further 9 gave the same response to both questions. Thus only 14 out of 34 respondents wanted a strictly larger reduction in the risk of death to compensate for a larger increase in what nearly all of the sample regarded as a worse injury. Moreover, of these 14, 5 gave answers which still translated into higher values of $U(X)$ than of $U(W)$.[12] So only 9 of the 34 individuals gave answers which actually generated higher values of $U(W)$ than $U(X)$.[13]

By contrast, there were very few such discrepancies observed in the SG sample: out of the 27 people whose responses were used in Table 2, only 3 gave answers which implied $U(X) > U(W)$. Although 13 others gave responses which implied $U(X) = U(W)$, 12 of these took the form of being unwilling to accept any risk of death as an alternative to the certainty of either $X$ or $W$ This latter does not appear to indicate misunderstanding, or 'protest' responses, since 6 of the 12 were willing to accept some risk of death as an

[12]For example, one of these five wanted a 20 in a million reduction in the risk of death to compensate for the 180 in a million increase in the risk of $X$, and a 15 in a million reduction in the risk of death to compensate for a 100 in a million increase in the risk of $W$ On this basis, $U(X) = (180 - 20)/180 = 0.889$, while $U(W) = (100 - 15)/100 = 0.85$

[13]A similar story can be told for injury $S$, where, despite the fact that 31 of the 34 respondents scored $S$ as strictly worse than $W$, only 15 gave RR answers which entailed $U(S) < U(W)$ – the aggregate result being that the mean utilities for $S$ and $W$ in Table 2 were the same – 0 78

alternative to the certainty of *S*. Rather, it may indicate that people were distinguishing between prognoses such as *R* and *S* which entail some permanent disability, and those such as *X* and *W*, where complete recovery and return to normal health is just a matter of time.

## V. CONCLUDING REMARKS

Although the study reported in this paper was commissioned by the DoT and therefore had specific policy objectives,[14] the results it yields are wide ranging and are applicable to any research involved in utility assessment. Despite the fact that the RR form of question is in principle arguably preferable on the grounds that it operates in the 'right' area of the Marschak–Machina triangle, the evidence reported here suggests that RR questions (at least in the form described above) are clearly vulnerable to confusion and/or bias. On the other hand, the SG questions appeared to be much more manageable, and not vulnerable to any obvious bias.[15]

However, we cannot claim confidently that SG questions necessarily are good proxies for the kinds of risk tradeoffs operating at the margin which more accurately characterise road safety measures. Although it is clear that the RR means and medians in Table 2 are an unreliable basis for assessing whether the slopes of indifference loci in this context do vary systematically as we move towards the top left-hand corner of the MM triangle, we still cannot discount that possibility.[16] It may be that indifference loci that 'fan in' provide the best descriptive model of human behaviour with respect to preferences over health and safety. Clearly this is an empirical question which requires additional research that will have implications for all economic appraisals where there exists uncertainty concerning outcomes.

Alternatively, it may be that a status quo effect biases responses so that it appears *as if* indifference loci fan in. Had we constructed the RR questions to involve movements in the same direction as the SG questions (i.e. 'What increase in the risk of death will exactly offset a given decrease in the risk of injury *I*?'), we would have controlled better for such possible effects. As it was, at the time when the pilot questionnaires were being constructed (1990), we were insufficiently alert to this possible source of discrepancy, and have had no opportunity to investigate it since. It is therefore one of the issues that we would suggest for inclusion in the agenda of further research.

What is required is some way of eliciting RR responses which allows genuine comparisons with responses to SG questions. There is no doubt that further research along these lines is needed: as things stand, the recommendations emerging from the main study conducted for the DoT were that little confidence could be placed in responses to the CV questions, and that the valuation of serious non-fatal injuries should therefore be based almost entirely upon the results of the SG questions. However, if behaviour in the top left-hand corner of the triangle does depart systematically from behaviour in the middle for reasons other than confusion, bias or status quo effects, the SG estimates would not be the most appropriate basis for policy. It would therefore be of considerable theoretical and practical interest to have a firmer indication of the nature and extent of any such departure.

## ACKNOWLEDGEMENTS

This paper reports results from a pilot study commissioned by the Transport Research Laboratory and the Department of Transport. We should like to thank those in TRL and DoT who helped us throughout. The opinions expressed in this paper are those of the authors, and do not necessarily represent the views of the TRL or the DoT.

## REFERENCES

Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R. and Schuman, H. (1993) Report of the NOAA Panel on Contingent Valuation, *Resources for the Future*, Washington DC.

Camerer, C (1995) Individual decision making, in *Handbook of Experimental Economics*, (Ed ) J. Kagel and A. Roth, Princeton University Press, Princeton.

Carson, R. T., Mitchell, R. C., Hanemann, W M., Kopp, R J., Presser, S and Ruud, P A (1992) A contingent valuation study of lost passive use values resulting from the *Exxon Valdez* oil spill, *Report to the Attorney General of the State of Alaska*, Westat Inc, Rockville, MD.

Chew, S H. and MacCrimmon, K. R. (1979) Alpha-nu choice theory: an axiomatization of expected utility, *UBC Faculty of Commerce, Working Paper 669*.

Dubourg, W. R , Jones-Lee, M. W and Loomes, G (1994) Imprecise preferences and the WTP-WTA disparity, *Journal of Risk and Uncertainty*, **9**, 115–33

Gafni, A. (1994) The standard gamble method. what is being measured and how is it interpreted?, *Health Services Research*, **29** (2), 207–24.

[14]In the light of the pilot study results, it was decided to mount a main study, involving a large representative national sample, using two elicitation procedures one of the CV formats not discussed here, and SG questions, along the lines of those discussed above

[15]Nor, incidentally, were there any obvious signs of bias in any particular direction in the subsequent main study (see Jones-Lee *et al*, 1993b).

[16]Though as noted earlier (see note 11), given linear indifference loci that all fan in, it is possible to place an upper bound on the extent to which SG responses will underestimate the required ratios of marginal rates of substitution. In particular, it can be shown that if the ratio estimated on the basis of the SG response is $\pi$ then the 'true' ratio can be at most $\pi/1 - \pi$. Thus, with $\pi$ equal, to, say 0.2 the SG response will underestimate the true ratio by at most 20% while with $\pi$ equal to 0.1 the underestimate will be at most 10%.

Gul, F (1991) A theory of disappointment aversion, *Econometrica*, **59**, 667–86.
Hausman, J. A. (ed.) (1993) *Contingent Valuation: A Critical Assessment*, North Holland, New York.
Jones-Lee, M. W. (1989a) *The Economics of Safety and Physical Risk*, Basil Blackwell, Oxford.
Jones-Lee, M. W. (1989b) *The value of avoidance of Non-Fatal Road Injuries, Report to the Department of Transport.*
Jones-Lee, M W., Loomes, G. and Philips, P. R. (1993a) Valuing the prevention of non-fatal road injuries: contingent valuation vs standard gambles, mimeo.
Jones-Lee, M. W., Loomes, G. O'Reilly, D. and Philips, P. R. (1993b) The value of preventing non-fatal road injuries: findings of a willingness-to-pay national sample survey, *TRL Contractor Report*, 330.
Kahneman, D., Knetsch, J. L. and Thaler, R (1990) Experimental tests of the endowment effect and the Coase theorem, *Journal of Political Economy*, **98**, 1325–48.
Kahneman, D. and Tversky, A (1979) Prospect theory: an analysis of decision under risk, *Econometrica*, **47**, 263–91.
Machina, M. J. (1982) 'Expected utility' analysis without the independence axiom, *Econometrica*, **50**, 277–323
Machina, M. J. (1987) Choice under uncertainty· problems solved and unsolved, *Journal of Economic Perspectives*, **1**, 121–54.
McDaniels, T. L. (1992) Reference points, loss aversion and contingent values for auto safety, *Journal of Risk and Uncertainty*, **5**, 187–200.
Road Accidents in Great Britain (1989), *The Casualty Report*, HMSO, London.
Viscusi, W. K , Magat, W. A., and Huber, J. (1987) An investigation of the rationality of consumer valuations of multiple health risks, *Rand Journal of Economics*, **18**, 465–79.
Viscusi, W. K., Magat, W. A. and Huber, J (1991) Pricing environmental health risks· survey assessments of risk-risk and risk-dollar trade-offs for chronic bronchitis, *Journal of Environmental Economics and Management*, **21**, 32–51

## APPENDIX A: INJURY DESCRIPTION CARDS

*H*

In hospital
- 1–4 weeks with burns/cuts
- Moderate to severe pain

After hospital
- Some pain/discomfort for several weeks, possibly months
- After healing no permanent loss of physical/mental abilities
- But prominent scarring permanently damages your appearance – you may feel self-conscious in some, possibly many, situations involving other people

*J*

Your normal state of health

*K*

Injuries resulting in death

*L*

In hospital
- Several weeks, possibly several months
- Head injuries resulting in severe permanent brain damage

After hospital
- Mental and physical abilities greatly reduced for the rest of your life
- Dependent on others for many physical needs, including feeding and toileting

*N*

In hospital
- Several weeks, possibly several months
- Loss of use of legs and possibly other limbs due to paralysis and/or amputation

After hospital
- Confined to a wheelchair for the rest of your life
- Dependent on others for many physical needs, including dressing and toileting

*R*

In hospital
- Several weeks, possibly several months
- Moderate to severe pain

After hospital
- Continuing pain/discomfort for the rest of your life, possibly requiring frequent medication
- Substantial and permanent restrictions to work and leisure activities

*S*

In hospital
- 1–4 weeks
- Moderate to severe pain

After hospital
- Some pain gradually reducing, but may recur when you take part in some activities
- Some restrictions to leisure and possibly some work activities for the rest of your life

*T*

- Sudden jarring/'whiplash' injury which does not require admission to hospital

After effects
- Frequent pain/discomfort in neck and/or back
- Some restrictions to work and/or leisure activities due to neck/back pain
- After 1–3 years, return to normal health with no permanent disability

*W*

In hospital
- 2–7 days
- Slight to moderate pain

After hospital
- Some pain/discomfort for several weeks
- Some restrictions to work and/or leisure activities for several weeks/ months
- After 3–4 months, return to normal health with no permanent disability

*X*

In hospital
- 1–4 weeks
- Slight to moderate pain

After hospital
- Some pain/discomfort, gradually reducing
- Some restrictions to work and leisure activities, steadily reducing
- After 1–3 years, return to normal health with no permanent disability

## APPENDIX B: QUESTION FORMATS

(i) *An Example of an RR Question*

Here are two types of injuries that you saw earlier.
Could you please glance at these two descriptions, to refresh your memory
The present risk of each of them is shown underneath:
80 in a million each year for *K* and 120 in a million each year for *R*

Suppose a particular road safety feature would reduce your risk of *K* but only by increasing your risk of *R* by 60 in a million.
I'd like to know how much you would want your risk of *K* to be reduced to make up for increasing your risk of *R* by 60 in a million.

(i) Increase in risk *R* definitely made up by reduction in risk *K* by ________ in a million.
(ii) Increase in risk *R* definitely not made up by reduction in risk *K* by ________ in a million.

You've said that reducing your risk of *K* by ________ (i) in a million would definitely make up for increasing your risk of *R* by 60 in a million, but that reducing your risk of *K* by ________ (ii) in a million would definitely not make up for it. If you had to give one figure between these two, what would your best estimate of the reduction in your risk of K which would exactly make up for increasing your risk of *R* by 60 in a million?

(ii) *An example of an SG Question*


Suppose you were in a road accident and suffered the injuries shown on *R*. However, you are told by the hospital that a special treatment is available which, if it succeeds, will put you in condition *J*. But there is a chance that the treatment could fail and if so you will die/suffer the disabilities shown on card *K*. You have to decide whether to have the treatment or not.

(i) Highest chance of failure at which definitely have treatment ________ in ________
(ii) Chance of failure at which definitely not have treatment ________ in ________

You've said that you would definitely have the treatment if there were ________ chances in ________ (i) of the treatment failing, but definitely not have the treatment if there were ________ chances in ________ (ii) of failure. If you had to give one answer, what chances between these two would make it most difficult for you to decide whether or not to have the treatment?