Valuing health and safety controls

A literature review

Prepared by
Department of Economics,
School of Psychology and
Experimental Psychology
for the Health and Safety Executive

CONTRACT RESEARCH REPORT
171/1998
Valuing health and safety controls
A literature review

Sue Chilton BSc PhD, Michael Jones-Lee BEng DPhil, Graham Loomes BA MSc and Angela Robinson BA MSc
Department of Economics
University of Newcastle-upon-Tyne
Newcastle NE1 7RU

Lorraine Hopkins BSc MSc and Nick Pidgeon BA PhD
School of Psychology
University of Wales
Bangor
Gwynedd LL57 2DG

Richard Cookson BA MPhil, Judith Covey BSc PhD and Anne Spencer BSc MPhil
Department of Economics
University of York
York Y01 5DD

Jane Beattie BSc PhD
Experimental Psychology
University of Sussex
Brighton BN1 9QN

This review considers the existing body of social science research on the valuation of benefits of health and safety controls. It assesses the strengths and limitations of that literature and identifies the most promising lines for future research.

The review takes as its starting point the fundamental prescriptive premise of conventional welfare economic theory, namely that public sector allocative and regulatory decision should, so far as possible, reflect the preferences, and more particularly the strength of preferences, of those who will be affected by the decisions concerned. From this it explores the argument that the value of any health and safety benefit to an individual might be a function of two things: the size of the risk reduction in that individual's probability of suffering harm; and the severity of the harm, judged according to that individual's personal tastes and preferences. It concludes that in order to actually deliver a method of valuing health and safety benefits which can be incorporated into public policy making, there is a need to consider first how far this really is the appropriate basic model to be working with and, secondly, how far it can be implemented in practice. It does so by looking at how economists, psychologists and researchers from other disciplines have addressed different aspects of these questions, with a view to describing and assessing the present state of knowledge, and identifying the most promising methods and directions for future research.

Sadly, Jane Beattie died in March 1997. Correspondence relating to this review should be addressed to either Michael Jones-Lee or to Nick Pidgeon.

This report and the work it describes were funded by the Health and Safety Executive (HSE). Its contents, including any opinions and/or conclusions expressed, are those of the authors alone and do not necessarily reflect HSE policy.
CONTENTS

CHAPTER 1 INTRODUCTION 1
  1.1 Transferability 3
  1.2 Knowledge and Judgement 4

CHAPTER 2 THE ECONOMICS LITERATURE 5
  2.1 Cost-Benefit Analysis 5
  2.2 Methods of Valuing Safety 6

CHAPTER 3 THE PSYCHOLOGICAL LITERATURE 9
  3.1 Definitional Issues 9
  3.2 The Psychology of Probability Judgement and Choice 10
  3.3 Psychometric Findings on the “Qualities” of Hazards 15
  3.4 Social and Cultural Issues 19

CHAPTER 4 CROSSING DISCIPLINARY BOUNDARIES 23
  4.1 The Challenges To The Economists’ Approach 23
  4.2 The Challenges To The Psychologists’ Approach 31
  4.3 Bridging the Gap: Valuing the “Qualities” of Hazards 37
  4.4 Constructive Approaches to Valuation 39

ACKNOWLEDGEMENTS 42

REFERENCES 43

APPENDIX A Summaries of Selected Publications on Contingent Valuation (CV) and other Direct Preference-Elicitation Methods in the Fields of Health and Safety (by Michael W. Jones-Lee). 59 - 135
CHAPTER 1 INTRODUCTION

When public sector bodies such as government departments or their agencies engage in activities (e.g. direct expenditure of public money and/or such things as legislation, regulation and guidance) which have implications for the health and safety of members of the population, how should those health and safety implications be valued and weighed alongside the other costs and benefits of those activities?

That is the central question we shall be addressing in this review, and in the sections below we shall consider the existing body of social science research relevant to that question, assess the strengths and limitations of that literature, and identify what seem to us to be the most promising lines of enquiry for future research to pursue.

We take as our starting point a fundamental prescriptive premise of conventional welfare economic theory, namely that public sector allocative and regulatory decisions should, so far as possible, reflect the preferences - and more particularly, the strength of preferences - of those who will be affected by the decisions concerned.

To see how such a prescription might be operationalised, we begin with a stylised portrait of a "typical" individual member of the population. Initially, this portrait may appear more like a caricature than a life drawing, but the reader is asked (temporarily) to suspend disbelief to some extent in the interests of laying bare the essential underpinnings of the current approach to valuing health and safety benefits, and setting the scene for the discussion that ensues.

Our typical individual is as follows. She has some level of current and expected future income and wealth and a set of personal values and preferences which, given the prices of goods and services available, result in a particular pattern of planned present and future consumption. Assuming that she knows her personal preferences quite precisely, and that she has good information about the qualities and prices of the alternatives available to her, the conventional assumption is that she chooses her pattern of consumption optimally, in the sense that there is no adjustment she could make which would give her greater well-being (or, in economists' terminology, utility).

Now suppose that some new safety device becomes available. Our stylised individual considers what benefit this safety device offers her in the form of some reduction in the probability of loss of quality and/or length of life. If this benefit has positive value for her, she will be willing to adjust her other (present and/or future) consumption in order to release some money from current income and/or savings to purchase the safety device. It is assumed that she will be willing to make such adjustments up to, but not beyond, the point at which the benefit gained from acquiring the safety device is exactly offset by the loss of utility from foregoing some amount of "other consumption".

At the point where the loss of utility from foregoing this other consumption exactly offsets the utility gained from acquiring the safety device, the money value of that foregone consumption gives us the individual's maximum willingness to pay (WTP) for the safety benefits under consideration. In other words, if the safety device were to cost more than this amount, she would not buy it, since the utility gained from acquiring it would be less than the utility of the other consumption she would have to forego. On the other hand, if the device cost anything less than her maximum WTP, she could (and should) acquire it by giving up some other consumption, since the utility foregone would be less than the utility gained from the reduction in risk to her quality and/or quantity of life.
Thus, on the basis of the above scenario, an individual’s maximum WTP provides a measure of the value to that individual of a given safety improvement. Two things appear to follow immediately from this.

First, if two safety devices - call them X and Y - both offer the same reduction in probability of harm, but if the severity of the harm addressed by X is greater than the severity of the harm addressed by Y, we should normally expect each individual’s WTP for X to be higher than their WTP for Y to an extent which reflects their personal evaluation of the relative degrees of harm involved. So, for example, if someone considers that permanent facial scarring and the loss of sight in one eye is a much more severe consequence than a simple fracture of one leg, and if safety device X will reduce their risk of the former by the same amount that safety device Y will reduce their risk of the latter, they might reasonably be expected to be prepared to forego substantially more other consumption to acquire X than to acquire Y, so that the relativities between their two WTP figures would reflect their personal evaluation of the relative severities of the two injuries upon their welfare.

The second conclusion which appears to follow from our stylised case is that if safety device X offers a bigger reduction in the probability of a particular injury than safety device Y, we should normally expect the individual’s WTP for X to be higher than their WTP for Y; moreover (although this may not appear to follow quite so obviously) as long as the changes in probabilities are relatively small and involve WTP values that are only a small fraction of the individual’s total wealth, we should normally expect that if safety device X offers a reduction in risk which is, say, twice as big as the reduction offered by Y, then the individual’s WTP for X should be very nearly double their WTP for Y.

A simplistic view might therefore take the value of any health and safety benefit to an individual to be a function of two things: the size of the reduction in that individual’s probability of suffering harm; and the severity of the harm, judged according to that individual’s personal tastes and preferences.

This in its turn may suggest a very appealing general method for calculating the value of health and safety benefits: simply construct a set of descriptions of health states representing different degrees of harm, and ask a representative sample of the population for their WTP to reduce their risks of injuries which would lead to each of these health states. Besides directly establishing a “tariff” of values for reducing the risks of ending up in the particular health states studied, values relating to other health states could be estimated by finding where people would locate these “new” states in the spectrum of health states already directly valued. Indeed, this is the kind of approach that lies behind much of the recent research into the measurement and valuation of health that has been commissioned, and endorsed to some extent, by the Department of Health (Dolan, Gudex, Kind and Williams, 1995; Dept of Health, 1995).

In other words, if we assume:

a) that people can reasonably easily and accurately envisage how their welfare would be affected by various degrees of loss of quality and/or quantity of life;
b) that they can handle the relatively small (reductions in the) probabilities of sustaining such losses;

c) and that they have sufficiently precise knowledge of their personal preferences to be able to judge the value of foregone “other consumption” which would just offset the utility of any particular safety benefit;

then we could derive a set of values which are a function of two dimensions - the size of the risk reduction, and the severity of the physical consequences of harm. Being a function of just these two dimensions, such values could then be applied across the very wide range of human activities where health and safety improvements might be achieved.
However, as stated at the outset, what we have above is a stylised case, a basic “model” which abstracts to some degree from the greater complexities of the real world in order to focus upon certain central concepts and principles. In order to actually deliver a method of valuing health and safety benefits which can be incorporated into public policy making, we need to consider a) how far this really is the appropriate basic model to be working with, and b) how far it can be implemented in practice.

Much of the rest of this review will discuss how economists, psychologists and researchers from other disciplines have addressed different aspects of these questions, with a view to describing and assessing the present state of knowledge, and identifying the most promising methods and directions for future research.

Without pre-empting that discussion too much, let us briefly outline the main issues involved.

1.1 TRANSFERABILITY

There are two principal issues here, which we might call transferability between contexts and transferability between private and public values. Let us take each in turn.

First, is the value an individual places on reducing the risk of a given harm by a given amount simply a function of the size of the risk reduction and the severity of the harm? Or are there other considerations - which we shall call contextual factors - which may also enter into the valuation? For example, is the value an individual places on a given reduction in the risk of facial scarring and the loss of sight in one eye independent of the circumstances under which such injuries might be sustained? Or might an individual’s WTP to reduce the risk of a particular set of injuries occurring as the result of a road accident be different from that same individual’s WTP to reduce the risk of the same set of injuries occurring as the result of a criminal assault? Or as the result of an accident at work? Or in the course of playing a contact sport? And so on.

Even within the terms of our simplified, stylised model, there is actually nothing which rules out the possibility that contextual factors may enter into individual tastes and preferences in such a way as to cause individuals’ WTP for a given reduction in the risk of a given loss of health status to vary, possibly quite substantially, from one set of circumstances to another. On the other hand, to date there has been little serious or systematic attempt to study this possibility. If there are variations, how substantial are they? And how far can we identify types of factors which tend to push the values systematically in one direction or the other?

The second issue concerns the extent to which estimates of the value of some public safety project can be obtained by aggregating the responses of the individual members of the affected population, each of whom (or, in practice, a representative sample of whom) is asked to state their WTP for the private benefit s/he would derive from that project.

To illustrate the point, consider first a private safety device which, if fitted to a car, would reduce the risk of death of the driver of that car by 1 in 100,000 per year. After inspecting his/her own preferences, each individual driver arrives at their personal annual WTP for this private safety device. If 100,000 such devices were to be fitted, we should expect that, on average, 1 fatality per year would be prevented as a result. By adding together all of the individual WTP responses, we would obtain a figure which represents a money measure of the aggregate personal welfare benefit of improving 100,000 individuals’ safety to the extent which is expected to prevent 1 fatality per year.

But now consider a public safety project - some scheme of road improvements, perhaps - which also affects 100,000 drivers and which is also expected, on average, to prevent 1
fatality per year. Can the aggregate value obtained in the previous paragraph be used to value the benefit of this public roadworks scheme? Or might individuals’ WTP for such a scheme reflect not only the value of the personal reduction in the risk of death which the scheme offers them, but also the value(s) they may place upon other considerations, such as the reductions in risk to other people, the equality (or otherwise) of the distribution of those reductions, the degree of voluntariness, control, personal responsibility, etc.? And if so, to what extent is it desirable and/or feasible to take account of - or perhaps, discount - the influence of these (and a number of other) considerations which have been shown by psychologists to be potentially influential in shaping people’s perceptions and attitudes to the acceptability of different kinds of risks?

1.2 KNOWLEDGE AND JUDGEMENT

The issues here may also be grouped under two headings: knowledge and judgement concerning information; and knowledge and judgement concerning values and preferences.

Our earlier assumption was that the typical individual had good information about all the available goods and services, including their qualities and prices; and it was implicitly assumed that this information would be processed efficiently and in a manner free of inconsistencies and biases.

In practice, of course, this assumption will often not hold. Many elements that enter into people’s consumption, and hence into their welfare, are not transacted in the market; and for many other goods and services that are marketed, individuals’ knowledge about their qualities may be really quite limited. Moreover, even when information about the qualities of goods and services - including any risks associated with them - is available, it may itself be based to a degree upon subjective judgements of “experts”, and there may not only be disagreements between experts on these matters, but also disparities between expert opinions and lay perceptions. In addition, individuals may experience a range of difficulties assimilating and processing any available information, and may be vulnerable to a variety of errors and biases.

All this might not be such a serious problem if these effects tended to cancel each other out in the aggregate - that is, if different experts’ estimates tended to be distributed more or less symmetrically around the “true” mean, or if differences in perceptions between some people in one direction tended to be balanced by opposite differences between other people, and so on. But as we shall see, there are doubts about how far this is actually the case. And even if people could be presented with uncontested and/or objective information, and could process it in ways that were free of systematic biases, there would still be the question of how far they could combine this information with sufficiently well-articulated and precise personal preferences in order to generate WTP responses which provide a reliable representation of their “true” values.

In the sections that follow, the issues outlined above will feature prominently. We shall organise the material in a way that largely mirrors the historical development of research in this field. In fact, for a number of years, researchers from different disciplines tended to plough their own furrows, largely in ignorance of - or at least, uninfluenced by - the lines of research being developed in adjacent fields. Thus Section 2 will provide a review of the “standard” approach employed by economists, largely without reference to any psychological or sociological research; while Section 3 will review the work done by psychologists and other social scientists, largely without reference to the concerns or activities of economists. Then in Section 4 we shall try to indicate how, in the face of difficulties or shortcomings encountered within the different paradigms, work which to some degree crossed disciplinary boundaries began to develop. The questions raised by that work, the limitations that remain, and the scope for further work will be discussed.
CHAPTER 2 THE ECONOMICS LITERATURE

2.1 COST-BENEFIT ANALYSIS

To repeat the point made in the Introduction, the fundamental prescriptive premise of conventional social cost-benefit analysis is that public sector allocative and regulatory decisions should, so far as is possible, reflect the preferences - and more particularly, the strength of preference - of those who will be affected by the decision concerned. Thus one requires a measure of strength of preference; and it would seem that such a measure is naturally provided by individual willingness to pay for desirable goods or services (and willingness to accept compensation for detrimental effects). Furthermore, the maximum amount that an individual would be willing to pay for a good or service is assumed to reflect not only the person’s valuation of the good or service relative to other potential objects of expenditure, but also the individual’s ability to pay - which is itself a manifestation of society’s overall resource constraint.

There then remains the question of how individual willingness to pay should be aggregated to arrive at an overall benefit measure for a desirable change. In conventional social cost-benefit analysis the aggregation is typically carried out on an unweighted basis: that is, the amounts that individuals are willing to pay are simply added up across all those affected by the allocative or regulatory decision in question - the resultant sum being a clear reflection of what the desirable change is "worth" to the affected group, relative to alternative ways in which each individual in the group might have spent his or her limited income.

The most obvious objection to this aggregation procedure is that it effectively replaces the principle of "one person, one vote" with that of "one pound, one vote" and, as such, inevitably accords the rich more "votes" in the allocative or regulatory decision process than the poor. There are, broadly speaking, three ways in which advocates of cost-benefit analysis have responded to this criticism, namely:

1. If one takes the view that income and wealth are already optimally distributed, then it is quite appropriate that those with greater command over resources should have a potentially more significant influence on the way those resources are allocated.

2. If, by contrast, one takes the view that income and wealth are not optimally distributed, then in principle the appropriate way to effect a redistribution is not by tampering with the results of cost-benefit analysis (by, for example, giving heavier weight to the willingness to pay of the poor) but, rather, by direct redistributional measures using taxes and transfers.

3. However, if for whatever reason, redistributional taxes and transfers are infeasible, individual willingness to pay can, if desired, by adjusted using "distributional weights" which, for those of more egalitarian persuasion, will tend to be inversely related to income or wealth.

Nonetheless, whichever of these stances is adopted, it is important to appreciate that the fundamental input to cost-benefit analysis is, in all cases, individual willingness to pay. A conventional analysis will simply add this up over all affected individuals and treat the result as the relevant measure, whereas an analysis with explicit distributional weights will clearly employ a weighted aggregate. It should be stressed that if distributional weights are to be employed then the levels at which these weights are set is essentially a matter of judgement for the decision making agency concerned.

In the case of safety effects, implementation of the willingness to pay approach relies upon the concept of the saving of a "statistical" life or the prevention of a "statistical" injury.
Recalling the example used in the Introduction, suppose that a particular safety improvement affords each of 100,000 individuals a 10^{-5} reduction in the probability of death during the coming year. This safety improvement will clearly reduce the mathematical expectation of loss of life by precisely one for the period concerned and is therefore referred to as the prevention of one statistical fatality or the saving of one statistical life. Aggregate individual willingness to pay for the safety improvement is then naturally referred to as the "value of saving one statistical life", or more succinctly as the "value of statistical life" (VOSL). A similar approach can be used to derive a value for the prevention of a non-fatal injury.

Finally, at the risk of labouring the point, it should be stressed that a WTP-based value of statistical life (or prevention of non-fatal injury) is an aggregate of many individuals' willingness to pay for (typically small) reductions in individual risk and is emphatically not the amount that any particular individual would pay to avoid the certainty of his or her death or injury.

That, then, is the underlying principle. The next question is: how can appropriate values be derived in practice?

2.2 METHODS OF VALUING SAFETY

There are basically two ways of obtaining empirical estimates of individual willingness to pay for safety which have found favour with economists: the revealed preference approach and the questionnaire approach.

Revealed preference methods attempt to infer the implicit trade-off that individuals make when faced with real economic decisions involving different levels of safety. Researchers have tended to concentrate on labour market decisions, since detailed data are available to estimate the implicit wage premium paid to workers for accepting a higher risk of death. Arguably, wage-risk studies of this kind can provide a useful estimate of the value of statistical life in the area of occupational safety. In most other areas of safety, however, and also in the case of non-fatal occupational injury and illness, it is extremely difficult - if not impossible - to obtain sufficient data to disentangle the various factors besides safety which may affect behaviour.

By contrast, the questionnaire approach appears to offer the possibility of collecting detailed data about precisely those safety effects that are of interest. The potential drawback with questionnaire methods, of course, lies not with the quantity of the data but with its quality. So the main aim of the early value elicitation literature was to find out whether or not responses to hypothetical questions could deliver sensible valuation estimates. Almost all of the early questionnaire studies in the field of safety used the approach known as "contingent valuation" (CV), which asks people quite directly how much they are willing to pay for a specified reduction in risk (or, less often, willing to accept for a specified increase in risk). These studies tended to focus primarily on the more "familiar" risks of immediate death, in areas such as road safety, fire safety and occupational safety. Despite considerable unexplained variation within and between different studies, it was felt that the questionnaire approach could deliver estimates which were at least of the same order of magnitude as estimates from revealed preference studies.

Until recently, the main item on the safety agenda was simply to find a value of statistical life. Values for preventing injuries were not a major focus of early work in this area. Moreover, although it was not explicitly argued that there was a single, universally transferable value of statistical life, little attention was given to considering factors that might suggest the use of different VOSL’s in different circumstances. However, as we shall see, work at the interface between economics and psychology has opened up this question by alerting researchers to the
possibility that a whole range of social and psychological factors may play a far more significant role in people's preferences than economists had initially assumed.

But before considering those factors, and other related issues, in more detail, we briefly review some examples of the standard CV approach in the more familiar areas of safety. Not only are these important areas of public policy in their own right, but it might be supposed that they are also relatively easy to value using questionnaire methods. One reason for thinking this is that people are used to making decisions about their own safety in these areas. In addition, the risks are relatively well-defined and "objective", since good statistical records are available.

An influential study in this area is the Jones-Lee, Hammerton and Philips (1985) survey of road safety, which employed professional interviewers and a nationally representative random sample. Since then, there have been some CV studies in the area of occupational safety and a number of CV studies of road safety in different OECD countries. There is evidence from these studies that CV responses to questions about familiar risks of immediate death vary, more or less, in the systematic way predicted by standard economic theory. In particular, we can identify three encouraging features of these CV estimates:

1. CV studies of different familiar risks of immediate death produce results that are of the same order of magnitude (e.g. Maclean, 1979, on fire safety; Jones-Lee et al., 1985, on road safety; and Gerking, De Haan and Schulze, 1988, on occupational safety).

2. CV studies of road safety in different OECD countries produce broadly similar results (e.g. Persson, 1989, in Sweden; Maier, Gerking and Weiss, 1989, in Austria; Miller and Guria, 1991, in New Zealand; Schwab-Christe, 1995, in Switzerland; and Desaigues and Rabl, 1995, in France).

3. Taken as a whole, CV estimates are similar to revealed preference estimates, although in the main they tend to be somewhat higher. This is confirmed by rough and ready comparisons between studies of "familiar" areas of safety - e.g. a wage-risk study of UK occupational safety by Marin and Psacharopoulos (1982) and a CV study of UK road safety by Jones-Lee et al. (1985) - and by a recent study which directly compared the two approaches in the same setting (Lanio, Pedro and Latour, 1995, a US study of occupational safety). This having been said, some more recent revealed preference wage-risk studies - using a simultaneous equation specification to reflect the fact that wage rates affect job risk (since wealthier people tend to choose safer jobs), as well as vice versa - have produced substantially higher estimates; see for example, Biddle and Zarkin (1988), Garen (1988) or Seibert and Wei (1992).

Although there are concerns that the bands of error on these valuations remain uncomfortably large, it can be argued that they appear to be no larger than those associated with scientific risk assessments of the facts about proposed health and safety controls. Thus there would seem to be a case for claiming that high quality CV studies can yield useful estimates of the value of statistical life in familiar areas of safety (Jones-Lee, 1989; Viscusi, 1992).

Against that, it has to be said that a number of doubts and reservations remain. For one thing, the statistical distribution of individual WTP responses is usually right skewed and widely dispersed, with the right skew sometimes being so strong that researchers feel compelled to use the median response as their estimate of aggregate WTP rather than the theoretically appropriate mean response (e.g. Viscusi et al., 1991). Another problem is that there are often a significant number of responses which appear so implausible as to suggest a misunderstanding of the question. Finally, doubts can be raised about the fact that certain key
variables are not always correlated with WTP responses in the way that standard economic theory predicts.

For example:

Income. Acton (1973) and Viscusi (1991) found no strong correlation between income and WTP. Most studies, however, have found that income is positively correlated with WTP, suggesting an income elasticity of about 0.3 (Jones-Lee et al, 1985; Persson, 1989; Maier et al, 1989; Miller and Guria, 1991).

Age. Persson (1989) and Viscusi (1991) found that WTP did not exhibit the "inverted-U" relationship with age which is predicted by standard theory and which has been found by most other studies which have examined this issue (e.g. Jones-Lee et al, 1985; Maier et al, 1989; Miller and Guria, 1991).

Baseline Risk. Smith and Desvousges (1987) found that WTP for a given magnitude of risk reduction decreased as the baseline risk increased, whereas standard theory predicts that, under normal circumstances, the reverse should be the case.

In addition, in the area where there have been many more CV studies than in the field of safety - namely, in valuing environmental goods and bads - there has been a growing body of evidence (about which we shall say more in Sections 2 and 3) to suggest that CV responses may be vulnerable to a number of systematic biases, inconsistencies and anomalies.

What may be thought of as the conventional economist’s reaction to this kind of evidence has been to suppose that such biases and inconsistencies can largely be attributed to shortcomings in the wording of questionnaires or some other aspect(s) of the study design - the implication being that such biases and inconsistencies can be greatly reduced, if not eliminated, by improved questionnaire/study design. Thus there have been some attempts to develop guidelines and/or establish more stringent criteria for “good practice” in CV exercises (see, for example, Cummings, Brookshire and Schulze, 1986; Fischhoff and Furby, 1988; Mitchell and Carson, 1989; Carson and Mitchell, 1995; Arrow, Solow, Portney et al, 1993). Current guidelines for a state-of-the-art questionnaire study would include: (i) the construction of realistic and credible scenarios, (ii) clear and unambiguous wording and presentation of information which has been developed and tested through extensive piloting and re-piloting, and (iii) various internal consistency checks. Indeed, in some cases, such as Arrow et al (1993), very specific recommendations (at least, in relation to obtaining environmental existence values) have emerged - namely, to conduct large-scale surveys using referendum-format dichotomous choice questions. However, it is far from obvious that problems with CV studies are rooted principally in defects in questionnaire/study design; nor is it clear that standardised guidelines/criteria would be equally applicable across domains as different as environmental existence values and health and safety benefits. Indeed, the literature from psychology and sociology might be thought to have rather different implications, and it is to this literature that we now turn.
CHAPTER 3 THE PSYCHOLOGICAL LITERATURE

The preceding Section presented the theoretical model underpinning contingent valuation in the domain of safety. Central to this approach is the assumption that people accurately evaluate (perceive) the risks that they will be paying to avoid/reduce. In this section we take a historical perspective on the psychological literature on how individuals appraise risks. Just as the early economic literature developed without much reference to research in other disciplines, so too the psychological research in this area proceeded largely in isolation from economic concerns: questions of how (or whether) risk perceptions should influence valuation were not pursued.

The goal of the psychological work was to ascertain how different risks are represented psychologically: both in terms of how accurately their quantity was represented with respect to some normative standard; and also which qualitative dimensions affected whether various risks were perceived as being similar to or different from each other. Beyond the psychological literature on individuals, we shall also consider other social science perspectives on risk perception, asking how group and societal factors influence risk perception: for example, whether all groups in society tend to represent risks similarly; and how social processes act to amplify certain risks while seeming to neglect others. The literature on the psychology and sociology of risk perception is very extensive. This section is not intended to give an exhaustive account, but rather to point the reader to the main lines of argumentation.

3.1 DEFINITIONAL ISSUES

In much of the formal safety regulation and risk assessment literature risk is typically defined either as the likelihood of some specified harm or else as some combination (very often the product) of likelihood and the severity of negative consequences. (see e.g. HSE 1995b, p7).

However, when discussing perception of risk, a number of other factors become relevant. A considerable body of psychological research has looked directly at the factors which influence individual judgements of subjective uncertainty in experimental settings (research which we briefly review below). More generally however we need to adopt a much wider concept than merely some subjective analogue of uncertainty, probability or expected loss.

In the 1992 Royal Society report risk perception was defined as ‘people’s beliefs, attitudes, judgements and feelings, as well as the wider cultural and social dispositions they adopt towards hazards (threats to people and the things they value) and their benefits’ (Royal Society, 1992, p89). This definition was drawn deliberately broadly because it is a wide range of multidimensional characteristics of hazards, rather than just an abstract expression of uncertainty and loss, which people appear to evaluate in forming perceptions. Furthermore, such evaluation may vary with the social or cultural group a person belongs to (depending for example upon their different value commitments), the historical context in which a particular hazard arises, and may also reflect aspects of the physical and human or organisational factors contributing to hazard, such as the trustworthiness of existing or proposed risk management.

The term “risk perception research” has therefore become a generic label to describe a very varied body of social science work on people’s responses to hazards, and which now includes input not just from psychology but sociology, anthropology, geography, decision theory, economics and policy studies. Othy notes (1990) that a more accurate definition of this diverse field would in fact be one of “hazard perception”. In practice, however, the label risk perception is typically adopted by researchers to describe their work.
The academic literature on risk perception continues to expand, and the interested reader is referred to recent reviews of the field in Rohrmann, Wiedemann and Stegelmann (1990), Royal Society (1992), Slovic (1992), Brun (1994) and Rohrmann (1995).

3.2 THE PSYCHOLOGY OF PROBABILITY JUDGEMENT AND CHOICE

3.2.1 Calibration of Experts and Novices in General

Research into risk assessment raises the question of whether people (both experts and novices) can accurately make judgements about risks, or whether there are systematic biases in their evaluations. For high-frequency cases where the outcome is well-defined, the accuracy of individuals’ judgements can be explored by relating the subjective probability for an event predicted by the individual to the actual outcome frequency on trials over the long run.

Consider, for example, the task of predicting the weather. The judgements of an individual at any particular level of probability (say 60% confident it will rain today) can be plotted against actual outcome frequencies. By taking an aggregate, for large enough samples of events and judgements, of all the occasions when the individual predicted there was a 60% chance of rain and then observing the actual outcome in these cases, this should lead, normatively, to 60% of these occasions actually resulting in rain. A similar exercise can be conducted for all occasions where an individual gives a different probability forecast, such as at the 70% confidence level, or at 90% etc. By plotting the level of judged probability at discrete levels of confidence against the actual outcome frequencies, a calibration curve is plotted, and when all points lie on the identity line the individual is said to be perfectly calibrated.

A large body of research now exists in this area of probability calibration (see Wright and Ayton, 1994). One regular finding is that novices in a particular prediction domain tend to be generally overconfident. That is when they predict certain events will each occur with 70% confidence, the actual outcome frequency is less than this across the set of events (e.g. Fischhoff and MacGregor, 1982). However, novices can also be well-calibrated on certain tasks (Fischhoff and Beyth, 1975). It is important to recognise here that this rather technical operational definition of under- or over- "confidence" from the probability calibration literature is rather different from its more common usage to reflect margins of error or ambiguity about a point estimate of probability.

One might expect experts to be uniformly well-calibrated, but this depends to a great extent upon the nature of the expertise domain. For example, while this is the case for professionals who can collect and analyse accurate feedback on their predictions, such as with the weather forecasting task discussed above (Murphy and Winkler, 1977), the picture is much bleaker in professions where feedback and accurate learning opportunities may be more difficult to come by. Hence, Christensen-Szalanski and Bushyhead (1981) found that physicians’ judgments that a patient had pneumonia were generally extremely overconfident, to the extent of being almost uncorrelated with the outcome probability! These results are important as they make the apparent divide between “objective” expertise and the "subjective" evaluations of laypeople much less clear cut.

3.2.2 Estimating Fatality Frequency in Particular

There is a considerable body of evidence concerning the estimation of the frequency of fatalities from common causes of death (vehicle accidents, cancer etc.). This shows that when people are asked about their understanding of how many fatalities arise as a result of an activity X, their rank ordered responses correspond by and large with those of available statistical estimates. This ranking is also consistent across different question formats (e.g. Fischhoff and MacGregor, 1983; Eiser and Hoepfner, 1991). Furthermore, there is some evidence (Brown and Green, 1981) that evaluations of personal safety are also reasonably
well correlated with statistical frequency: hence the personal risks of travelling by train are correctly seen as lower than those of riding a motorcycle or of rock climbing. These studies seem to support the idea that lay judgements of the rank ordering of frequencies of harm are reasonably accurate. This is a significant finding, because if people rank the “importance” of the risks of various activities in a different order from their ranking of the frequencies of fatalities associated with those activities, this cannot be attributed to a lack of information or awareness concerning those relative frequencies; and this in turn suggests that the notion of risk means something more to people than just expected fatalities.

Regarding absolute or quantitative judgements, the standard findings here, following the work of Lichtenstein, Slovic, Fischhoff, Layman and Combs (1978), are that when asked for frequency judgements in comparison to a specified anchor (such as the number of deaths from automobile accidents across the US) people tend to overestimate the prevalence of low frequency causes of death (botulism) and underestimate those of high frequency (heart disease). This finding suggests a regression effect in which people’s judgments are insufficiently discriminating, perhaps because of random error. However, in the next section we raise the much more problematic issue of systematic bias across individuals.

3.2.3 Heuristics and Biases in Probability Estimation

The issue of systematic bias was raised by the research programme of Kahneman and Tversky (see, e.g., Kahneman, Slovic and Tversky, 1982) in which they demonstrated a number of “heuristics” (mental shortcuts or rules-of-thumb) which people use in simplifying the task of estimating probabilities (among other quantitative tasks). Kahneman and Tversky proposed three main heuristics that would be expected to lead to systematic bias in risk estimation: availability, representativeness, and anchoring and adjustment. While argued to be generally adaptive (see Jungermann, 1986), these heuristics can in certain circumstances lead to large, systematic errors in probability assessment. These results are important because they call into question the assumption that individual errors will be random or unbiased, and will tend to cancel each other out, given large enough samples.

For example, the availability heuristic states that people judge the probability of an event in part by the ease with which it is accessed from memory. Slovic, Fischhoff and Lichtenstein (1979) found that vivid, imaginable causes of death (tornado) receive similar estimates to non-vivid ones (asthma) which occur with much higher frequency (in this case, by a factor of 20). Other studies of availability effects, including Fischhoff, Slovic and Lichtenstein (1978), suggest that experts evaluating probabilities of safety failure will pay insufficient attention to causes that are not elaborated in their mental model of the system. In the next Section, we shall see that availability has also been hypothesised to explain so-called “embedding effects” in contingent valuation.

Kahneman and Tversky (1972) have also argued that one of the most important rules of thumb used in forming probability judgements under many circumstances is the representativeness heuristic. Their original definition was as follows:

‘An event A will be judged a member of a category B (or a sample seen as typical of its parent population) to the extent that it is representative of that category or population; i.e. (a) similar in its essential properties to its parent population, and (b) reflects the salient features by which it is generated’ (1972, p431)

For example, in their early studies Kahneman and Tversky demonstrated that the sequence HHTHTH is regarded by many people as the more likely outcome of a sequence of six tosses of an unbiased coin than is the sequence HHHTTT, despite the fact that both are, normatively, equally likely outcomes. They argue that the former is more representative of, that is similar to or consistent with our expectations of, a “typical” random sequence (although see Lopes, 1982). Used as a cue to judge frequency (where a more representative match is taken to imply
a higher likelihood) it can sometimes lead to inappropriate judgements where similarity and frequency depend upon different properties of the event in question. Of the three original heuristics, representativeness has generated the greatest amount of debate and subsequent research, and is claimed to be implicated in a number of commonly researched biases of reasoning (Kahneman, Slovic and Tversky, 1982). These include such things as insensitivity to the impact of base rate probabilities in Bayesian inference (Bar-Hillel, 1980), and the so-called conjunction fallacy of probability judgement (Tversky and Kahneman, 1983; although for critical comments upon the representativeness explanation of this latter phenomenon see Fisk and Pidgeon, 1996).

The anchoring and adjustment heuristic (Tversky and Kahneman, 1974) results in absolute estimates being influenced up or down by the provision of different anchors (starting points), even if the subject can readily see that the anchor provides no information about the correct answer. In the case of risk estimation, final estimates will be correlated with the starting anchor, and are thus open to manipulation by self-interested parties or the media. In the next Section we shall see that there is also some evidence of starting point bias in the elicitation of willingness to pay responses.

3.2.4 Framing Effects

Starting point bias is one example of how the way in which a question is posed can systematically influence the answer given. Tversky and Kahneman (1981) have described many such “framing” effects, in which apparently irrelevant changes in wording of the question (from the perspective of normative theory) produce substantial shifts in response. The most famous example is the “Asian Disease Problem”, in which subjects must choose between a risky and a safe programme to combat the disease’s effects on the US population. The irrelevant wording change was to describe the effects either in terms of losses (lives lost) or gains (lives saved) - although the objective outcomes were the same in each case. People in the losses version made many more risk taking choices than those in the gains version. This effect of gains/losses on risk attitude has proved robust in a variety of risk decision problems (see Lopes, 1987), and has potentially serious consequences for attempts to elicit preferences for health and safety expenditure. The question is also raised of whether we can definitively state that one framing of a problem is a better resolution than any other (Bell, Raiffa and Tversky, 1988).

3.2.5 Choice Heuristics and Constructive Decision Processes

Since the 1950s Herbert Simon (1955, 1957) and other psychologists have argued that, in the face of complex choice situations, economically rational optimization is in general far too demanding a task for the individual decision maker. For example, Simon argues that:

'Because of the psychological limits of the organism (particularly with respect to computational and predictive ability) actual human rationality striving can at best be an extremely simplified approximation to the kind of global rationality which is implied, for example, by game theoretic models” (1955, p102).

He also argues that much human behaviour is better characterised by “satisficing”. Hence Simon questions both the descriptive and crucially the prescriptive validity of rational economic models in the domain of behavioural choice; the intrinsic calculation demands of rational economic choice being incompatible with human cognitive limitations and capacity. Simon rejected, as a practical description, the notion of the economic individual (in the sense that she maximizes global subjective expected utility, or some pseudo-variant of this). Rather, actual choice behaviour could be described in terms of the principle of bounded rationality. That is, in order to cope with the complexities of choice the individual constructs a simplified cognitive representation (or small-world) of the environment, and then chooses more or less rationally with respect to that representation.
This theoretical position, first taken up by Simon, provided much of the stimulus for the research program of Tversky and Kahneman, noted previously, on heuristics and biases of probability judgement and choice. And, more generally, many psychologists interested in behavioral decision making have, over the past 25 years, been extensively concerned with the identification of the particular shortcuts and mental rules of thumb used by individuals to simplify complex problems (e.g. Slovic, Fischhoff and Lichtenstein, 1977; Einhorn and Hogarth, 1981; Baron, 1988). Of course, many of the experimental paradigms used by these researchers present participants with very finely balanced choices - the early, and elegant, intransitivity studies of Tversky (1969) being a good case in point. What is clear, however, is that such departures from "rational choice" do not so much represent random error or noise about some ideal strategy, brought out only by the peculiarities of the task at hand, but rather stand as more fundamental demonstrations of an incompatibility between prescriptive (economics/statistical) and descriptive (psychological) models of the choice process.

An important tool in further exploring the dynamics of choice has been the use of process tracing techniques (think aloud protocols, eye-movement techniques, or information boards) to map the choice rules people do deploy when faced with multiattribute decision situations (Payne, Braunitz and Carroll, 1978; Harte and Koele, 1995). Early research in this area led to taxonomies of the simplifying rules (such as the conjunctive rule, disjunctive rule, maximin, and elimination by aspects; see e.g. Montgomery and Svenson, 1976; Svenson, 1979) that individuals would deploy in arriving at choice. An interesting and robust empirical result is that selection of simple choice rules is not a static process, but varies for any individual contingent upon changing features of the choice task (Payne, 1982). For example, as complexity increases by increasing the number of choice alternatives or attributes in a choice task there is more use of simple attribute-based heuristics, such as elimination by aspects (Svenson, 1979). Such rules ignore part of the information available but have the advantage of making choice more tractable. Such neglect of information can be said, in one sense, to be strictly sub-optimal. However, seen from the perspective of attaining a good enough choice (for example, judged by avoiding all very bad outcomes, and where small departures from the strictly optimal might not matter; see von Winterfeldt and Edwards, 1973) their use can be entirely reasonable.

Other researchers, following up on another of Simon's basic propositions, have been interested in the ways in which individuals construct or structure the small worlds within which choice heuristics can then be used (see e.g. Montgomery, 1983). And today's psychological models of choice fully recognise the fact that problem structuring and choice rule selection are closely related, often as dynamic and creative processes involving various sub-operations of search, editing, elaboration and evaluation. Perhaps the most developed work in this area is that by Payne, Bettman and Johnson (1993). They review a series of studies of actual human choice, as well as simulations of the accuracy of typical choice rules in different information environments (also Thorngate, 1980; Pidgeon, 1986). Payne et al argue that considerations of cognitive effort in making a choice, as well as desired level of accuracy, can help to explain how problem structuring and heuristic use are both flexible and often rationally adapted to the different environments in which preference and choice has to take place. The notion of preference and choice as "one-shot" operations that can be measured against more formal static prescriptive models has therefore been replaced with a more dynamic cognitive perspective; and with this the vision of a boundedly (ir)rational individual has made way, in the 1990's, to that of an adaptive, constructively rational one.

3.2.6 Group Effects

It is tempting to imagine that the effect of individual biases in probability estimation might be reduced by group decision making. Important expert decisions, for example, are often made by groups. However, the literature on group decision making suggests a number of group biases. Janis' (1982) work on retrospective analysis of foreign policy decision making
disasters shows that groups with insufficiently divergent views can become overconfident in courses of action that outsiders would regard as extreme, and Turner and Pidgeon (1997) describe many of the sociological causes of cognitive insularity in groups and organizations and their role in producing disasters (see also Vaughan, 1996, for a complementary analysis of NASA's flawed decision to launch the Space Shuttle Challenger). Other studies by Stoner (1968), and Einhorn, Hogarth, and Klemmer (1977) all provide evidence that group processes often appear to move group decisions to a more extreme position (sometimes more risk-taking, sometimes more conservative) than members initially held. The suggested solution is to improve collective problem-solving by ensuring that the group is comprised of individuals with differing orientations to the problem, operating in an atmosphere in which divergent opinions are valued and can be expressed (Janis 1982). In these latter circumstances, group estimations may be more successful than the average, or even the best, of the individual estimations.

3.2.7 Appraisal of Objectively Given Probabilities

The literature reviewed above largely examined how people assess risks from their own knowledge and experience. A different question is how they deal with probabilities supplied to them, for example by expert assessors. A variety of studies have shown that people do not always weigh objectively given probabilities adequately in decision making.

One problem lies in how people deal with very small probabilities. As we have seen from the literature on fatality frequency estimation, such probabilities are sometimes overestimated or overweighted. Other more artificial laboratory studies confirm this, but also show how sometimes small probabilities will be discounted altogether (Kahneman and Tversky, 1979). This issue raises particular problems for eliciting accurate WTP responses because many health risks (especially of fatality) are, by their nature, very infrequent events, and as yet there is no fully articulated theory of when task demands will lead either to overestimates or else underestimates of small probabilities. The reverse of this problem is that people may have a tendency to overweight probabilities of 1.0 (Kahneman and Tversky called this the "certainty effect"). In such cases, respondents are willing to pay disproportionately more to reduce a risk to zero (certainty of safety) than for other similar marginal risk reductions. (See Viscusi, Magat and Huber, 1987, for an example of this in the context of willingness to pay for reductions in the risks from domestic chemical products.)

It has also been shown repeatedly that people have an aversion to dealing with ambiguity in probabilities (that is, where one can at best only say that the probability of an event lies somewhere in a range between x and y). For example, Ellsberg (1961) showed that people are less willing to bet on the basis of ambiguous probabilities than on point estimates of the same mean value (see also Frisch and Baron, 1988). In many areas of risk assessment for health and safety there is great uncertainty about the point estimate (e.g. probability of a loss of coolant accident in a nuclear power plant). This may be due to disagreement between different experts' analyses of the problem, or, for example, due to uncertainties about the way that systems in particular states will operate or the unpredictable effects of human errors (Pidgeon, 1988; Turner and Pidgeon, 1997). Again, some of the perceived differences in risk (see below) may be due to differences in the ambiguity of the probabilities. It may also be that a high degree of ambiguity correlates with particularly dreaded hazards, and Johnson and Slovic (1995) discuss some of the difficulties of communicating ambiguous risks.
3.3 PSYCHOMETRIC FINDINGS ON THE "QUALITIES" OF HAZARDS

A somewhat separate and highly influential stream of psychological research has looked at notions of "risk" which are rather broader than probability or accident frequency per se.

The bulk of work has concentrated upon the observation that lay definitions of risk and its acceptability are fundamentally related to much wider qualitative characteristics of the event. For example, Starr (1969) first made the important distinction between voluntary and involuntary exposure to risk as a factor influencing acceptability; while Otway and von Winterfeldt (1982) suggest that the acceptance of technological risks (particularly that of nuclear power generation) may be influenced by a number of dimensions, including:

- Voluntariness of exposure
- Lack of personal control
- Uncertainty about probabilities or consequences of exposure
- Lack of personal experience of risk (fear of unknown)
- Difficulty in imagining risk exposure
- Effects of exposure delayed in time
- Genetic effects of exposure
- Infrequent but catastrophic events (or "Kill Size")
- Benefits not highly visible
- Inequity (benefits go to others/risks fall on us)
- Accidents by human failure rather than natural causes

The work of Green and Brown in the UK also makes the important distinction, for risks that are immediate in their effects, between perceived threats to personal safety, and threats more widely to society: for example the risks from rock climbing versus those from a major chemical plant (Green, 1979). Threats to personal safety are characterised primarily by the perceived chances of an accident and the likelihood that death will follow. Threat to society, on the other hand, is related more to consequences: that is, to the number of people who will die in a year, or the maximum number of deaths from one particularly bad incident (and this could be extended to include other damage, such as to the fabric of a society or community). In Green's work, threats to the individual and society are not strictly polar opposites, but rather represent two independent dimensions along which perceived risks are categorised. Hence rock climbing is seen as a threat to the individual but not to society, while car transportation is both an individual and a societal threat, with nuclear power generation being primarily a societal threat. Clearly, there is a link here to the more formal use of the notion of "societal" risk in UK safety regulation policy, which draws in part upon a belief in public aversion to particular types of large-scale hazards (Royal Society, 1983; HSE, 1992). It is clear, however, that several of the characteristics listed in the table above are closely correlated with, and would also distinguish in very broad terms between, such so-called individual and societal risks. The point to which the discussion returns later is which of these particular characteristics - kill size, (in)voluntariness, lack of control etc. - underlies any observed aversion to risk.

3.3.1 Basic Psychometric Findings

Perhaps the best known research in this area is the psychometric work of Fischhoff, Slovic and Lichtenstein in Oregon (see Slovic, 1987, 1992; Rohrmann, 1995). This research programme began in part as an exploration of Starr's (1969) early findings on voluntariness, but more generally of the "revealed preference" approach to valuing safety. These investigations began, as Slovic notes (1992), as an attempt to map the "personality of hazards". Typically the notion of risk is left undefined in such a study, so that an individual's own construals of the concept are elicited. Also commentng retrospectively, Fischhoff (1990) notes that the findings demonstrated (a) that the revealed preferences method was
fundamentally flawed as a reflection of public preferences (since accepted risk does not always mean acceptable risk), and (b) that the qualitative dimensions, along with risk and benefit ratings, were intercorrelated in rather subtle and complex ways.

As is now well known, this classic series of studies (Fischhoff, Slovic, Lichtenstein, Read and Combs, 1978; Slovic, Fischhoff and Lichtenstein, 1980) show that for general and technological hazards, three principal independent factors account for the major pattern of intercorrelations amongst people’s ratings of hazard along the principal dimensions of hazard quality listed above.

The factors identified by Fischhoff, Slovic and colleagues are as follows

**Factor 1: Dread Risk**

- this factor is typically found to be positively correlated with:
  
  lack of control over risk
  catastrophic/fatal consequences, or risk potency
  involuntary risk
  risk not easily reduced
  inequity in risk distribution
  threat to future generations

**Factor 2 - Unknown Risk (or Lack of Knowledge about Risk)**

- this factor is typically positively correlated with:
  
  not observable
  unknown (to those exposed or to science)
  new risk
  delayed impacts

**Factor 3 - Number of People Exposed**

The "factor space" represented by the first two of these dimensions, which account for most variation in the scores, is now widely reproduced (see e.g. HSE, 1995a). Furthermore, a group of follow-on studies, some with larger and more diverse samples than the original, have broadly replicated this pattern of at most two or three independent factors. The precise interpretation to be placed upon the factors are sensitive to aspects of the hazard set judged, the qualitative dimensions included in the study design, and the methods of data analysis adopted (for overviews of this work see Royal Society, 1992; Slovic, 1992; Rohrmann, 1995).

A number of significant points to note, however, which have emerged subsequent to the original studies are as follows.

Viewed singly, different hazards can have highly varied “qualitative profiles” for their perceived risks and benefits as found, for example, by Gardner and Gould (1989). This may imply that the impact upon risk evaluations of any particular hazard’s qualitative profile may have to be judged on a case-by-case basis, and be related to the particular context within which a hazard arises (Brun, 1994). Also related to this is the point raised by Savage (1991,
1993) that some hazards (e.g., cancer, nuclear power) may be highly idiosyncratic, or stand out as exceptional, in relation to others in the qualitative factor space.

The original work used as stimuli a general set of hazards. Studies utilising more specific hazard sets (medical risks, transportation risks - see Slovic, 1992) tend to evoke different factor spaces, although there may also be a "Russian Doll" effect. That is, for any generic hazard located within the main Dread/Unknown factor space (e.g., rail transportation) a subset of more specific activities can be constructed (passenger train, nuclear or chemical freight etc.) which are psychologically differentiated once again along the broader dimensions of Dread and Knowledge (Kraus and Slovic, 1988).

As well as the intercorrelations between the qualitative dimensions themselves (dread with accident size, unknown with new risk etc.), the hazard set will also influence the correlation to be found between risk/benefit judgements and particular qualitative dimensions. For example, Gregory and Mendelsohn (1993), in a re-analysis of the original Slovic et al (1980) data, report catastrophic potential (or size of accident) as negatively correlated with perceived risk, contrary to intuitive expectations. They attribute this to the inclusion in the design of a number of activities (e.g., smoking) that the respondents rated as highly risky but which are also seen as essentially "individual" harms (cf. the individual-societal risk distinction noted already). Again this cautions against simple generalising of any single theoretical interpretation across a range of hazards.

It is tempting to interpret the two principal dimensions of (a) Dread and (b) (Lack of) Knowledge as a direct mapping onto the more formal decision analytic concepts of "consequence" and "uncertainty". However, a complexity here concerns the loading and interpretation of the critical dimension "control" (or lack of it) over risk. In the Slovic et al (1980) study, lack of control is part of (i.e. positively correlated with) the "Dread" dimension - presumably because what we cannot control we often fear, and hence see as risky. However, seen from another perspective (Vlek and Cvetkovich, 1989; Teigen, 1994), one facet of the uncertainty associated with some loss or accident seems to hinge upon whether the hazard is controllable or not (individually for some risks and by society for others). The first Oregon study (Fischhoff et al, 1978), as well as some subsequent European studies, have found a two factor structure with control loading onto the second "Uncertainty" dimension (Vlek and Stallen, 1981; Teigen, Brun and Slovic, 1988; Brun, 1992; Puy and Aragones, 1994). This inconsistency may of course just reflect a methodological artifact. More probably it points to the fact that control is a complex and multifaceted concept in and of itself, and one not adequately captured by any single response scale, nor even any single study.

Finally, Savage (1993) makes the methodological point that while "Dread" is now well understood by risk professionals, it is not a concept people use readily in day-to-day discourse. He argues that a general "Unease" or "Insecurity" rating captures this aspect more closely, and is easier for respondents to use. Savage's revised scale ranges from "Feel Calm (when thinking about a hazard)" at one end to "Feel Nervous" at the other.

3.3.2 What does Most Work in Driving People's Perceptions of Risk?

A key question arising from the psychometric work is how the qualitative factors relate to evaluations of riskiness and its acceptability or tolerability. This is a complex issue, which again must be viewed to some extent upon a case-by-case basis. Brun (1994) points out that the qualitative dimensions may impact upon several different aspects of risk evaluation: perceived magnitude of risk and its benefits; the significance/importance of a hazard; judgements over who has responsibility for safety; willingness to pay for safety benefits (to which we return later); and the signal potential of an event (that is, what an accident portends).
However, in broad summary of the relationships between the qualitative factors and riskiness

1: One conclusion drawn from the original Oregon studies was that as one moves from the bottom left to the top right-hand quadrant in the factor space, a hazard will be seen as more risky, less tolerable (nuclear power was the prototypical example), more in need of regulation and control, and that an accident with such an activity or technology would be a significant portent of further things to come (the "accidents as signals" hypothesis; Slovic, 1987). In summary, there seems to be a particular association between extreme perceived riskiness/unacceptability and the more dread and unknown hazards.

2: It is clear from the Oregon studies and subsequent research (see Rohrmann, 1995) that of the three factor dimensions, it is the first one (Dread/Potency) which tends to be most consistently and positively correlated with directly expressed ratings of "gross" perceived risk, behavioural intentions (although intentions do not necessarily always predict behaviour), and attitudes towards tolerability or further reduction in risk. Other things being equal, the higher an activity's dread score the more risky it is seen to be, and the greater is the desire for control. However, it may be that Dread by itself, despite its predictive power across a set of general hazards, holds the status of an intermediate variable only - in that the hazard which makes us uneasy is then seen as risky, but what makes us uneasy will in turn differ from hazard to hazard and context to context (lack of control when flying; threat to future generations and catastrophic potential with nuclear power etc.; Gardner and Gould, 1989).

3: The direct relationship between the second psychological dimension (Knowledge) and perceived risk levels and acceptability judgements appears to be very much weaker that that of Dread. Knowledge about a risk is of course a complex and subtle point, to which the discussion returns later when we consider the issue of willingness to pay for safety controls.

4: It is also clear (Morgan, Fischhoff, Lave and Fischbeck, 1994) that the third factor, Number of Deaths and Injuries, also influences risk evaluations as might be expected. Likewise, benefits also impact upon judgements of tolerability in a broadly predictable way (more benefit makes an equivalent risk seem more tolerable and vice versa).

5: In a recent re-evaluation of the typical results, Alhakami and Slovic (1994) report perceived risk and benefit as inversely related, rather than independent dimensions (as might be expected on normative grounds). This is arguably because once an attitude is formed towards an activity, it may drive both risk and benefit evaluations, but in opposite directions.

6: Reviewing current UK risk regulation, a recent report on safety standards in government (HM Treasury, 1995) argues that the classic risk perception factor space does not map directly onto current regulation practice, in the sense that examples of heavily regulated risks can be found in all quadrants of the traditional Dread/Knowledge factor space. Concentrating upon the dimensions of control and involuntariness alone, however, the report argues that government regulation is very often present where a risk is not controlled by those exposed to it, and even more so where exposure arises involuntarily. Bohnenblust and Slovic (1996) develop some similar lines of argument, and a decision making framework - linking these issues of perception to explicit tolerability weightings - which has been successfully applied in continental European safety regulation and engineering practice for a number of years now. This underlines the need to develop a contextually sensitive and coherent theoretical analysis of the control issue, and its importance for both risk perceptions and safety management, as the comments upon risk and trust, below, also highlight.
Finally, in scientific terms, there is a question mark against the precise theoretical interpretation to be placed upon the psychometric findings. They undoubtedly provide us with an impressive empirical data base, and show us that individual conceptions of risk are extraordinarily rich and complex in their variety and detail. However, some have argued (Royal Society, 1992; Kasper, 1992) that the psychometric research would benefit from theoretical marriage with certain of the concepts and theories of the social and cultural approaches reviewed in the next section, and some progress is now being made here (Marrs, Langford and O’Riordan, 1996; Peters and Slovic, 1996; Earle and Cvetkovich, 1995).

3.4 SOCIAL AND CULTURAL ISSUES

There has been a growing awareness within the social sciences that social, cultural and institutional processes may have a major role in risk perception (see Krimsky and Golding, 1992; Bayerisch Rück, 1993). The perceiver of risk is rarely an isolated individual, but is a “social being” defined through a variety of group (e.g. family, friends), institutional (place of work), and cultural (ethnic, national) relationships. And whereas the 1983 Royal Society report on risk did not review this topic in any depth, an extensive part of the 1992 revision was devoted to this issue. We therefore consider some of the main points raised both there and since.

3.4.1 Attitudes and “World-views” Towards Risk

A criticism of the early Oregon studies on risk perception was the relative insensitivity of the study designs to the analysis of group differences. One early psychometric study, using a relatively large representative sample in the Netherlands (Vlek and Stallen, 1981) is an exception. These authors found evidence of differences in the way their subjects viewed the (existing) degree of organised safety associated with a hazard. Half of their sample viewed the existence of a high degree of organised safety (as with an industrial hazard process plant) as implying higher risk than did the more individual hazards (smoking). For the other half of the sample the reverse relationship was found. This might be interpreted in terms of differing attitudes of trust (or conversely distrust) regarding the existing institutional arrangements for handling hazards. Furthermore, Vlek and Stallen concluded that “the use of group average ratings makes far less sense than is often believed” (1981, p269).

If this is so, and risk perceptions are strongly related to group differences and attitudes, then we should expect to find significant correlations between risk evaluations and measures of general attitudes (for example, regarding fundamental ethical values, lifestyle, technology or the environment), or with sociodemographic variables (age, socioeconomic status, gender, political affiliations etc.). The evidence, however, while suggestive of some weak group differences, is far from clear cut.

Social psychological studies looking at people’s attitudes (positive or negative) towards nuclear power (for a review see van der Pligt, 1992), or towards the relationship between technology and the environment (Cotgrove, 1982; Buss, Craik and Dake, 1986) show that prior orientation is indeed often related to the aspects of the problem which are most salient for people (see also Alhakami and Slovic, 1994). The picture concerning value-orientation is, however, probably more complex than any simple categorization of people as “pro- or anti-” can capture. For example, consider attitudes to technology: Gardner and Gould (1989) found that attitudes favouring stricter risk regulation often go hand in hand with the desire to see a technology more widely developed (suggesting in turn that public calls for stricter regulation cannot be lightly brushed aside as solely representing an anti-technology bias).
Likewise the four way sociological typology (hierarchist, egalitarian, individualist, fatalist) posited by the so-called cultural theory of risk (Douglas and Wildavsky, 1982; Adams, 1995) does not easily reduce to a set of “world-view” measures. Reflecting this, in a general review and empirical investigation of this question, Sjöberg (1997) concludes that while significant differences can sometimes be found (for example, between fundamental value orientation and attitudes to technology, or between gender and risk appraisal, as noted above) correlations are usually very weak (most $r < 0.3$) and therefore explain only a very small amount (typically less than 10%) of the variation in perceived risk scores. Sjöberg also notes that attempts to utilise quantitative measures of the four biases have met with mixed success.

However, while it is now clear that individuals cannot easily be categorised in terms of the four world-views using standard quantitative techniques (e.g. Dake, 1991), this does not rule out empirical support for some variants of cultural theory, particularly where in-depth qualitative research of particular social groups is combined with more quantitative measures (see the extensive discussion of this issue in Marris, Langford and O’Riordan, 1996). Such a conclusion would be consistent with a view that people’s value orientations are not so much stable individual traits but labile and constructed from a set of available societal discourses about such things as danger, blame, trust and accountability (Kemp, 1990; Rayner, 1992; Pidgeon, 1995). The question then becomes less one of who holds to what attitude, but of which discourses are available (and why) to be invoked in any particular context or conflict, or to contribute to social amplification effects (Kasperson, 1992) on the occasions when things do go badly wrong.

3.4.2 Socio-demographic Variables

Of the general socio-demographic variables that have been studied there is some consistent evidence that gender (Flynn, Slovic and Mertz, 1994; Burke, Jenkins-Smith and Slovic, 1994) may influence evaluation of risk: where there is a significant difference, women tend to give higher risk ratings and see more threat to the environment (although in a recent study by Greenberg and Schneider, 1995, this difference disappears for respondents where the immediate neighbourhood has severe levels of technological hazards or other stressors). Age (Fischer, Morgan, Fischhoff, Nair and Lave, 1991), occupational affiliation (Sjöberg and Sjöberg, 1991; Jenkins-Smith and Bassett, 1994) and ethnic group membership (Vaughan and Nordenstam 1991; Vaughan, 1995) also influence risk perceptions. On the other hand, in their large US study, Gardner and Gould (1989) report very little relationship between sociodemographic variables and risk perceptions. Like value orientation, the demographic influences (at the level of correlates across individuals) are typically weak.

3.4.3 Risk and Trust

A very recent set of arguments and studies have focussed upon the relationship between risk and trust (and in particular the institutional arrangements for handling risk; Royal Society, 1992). The question of trust and risk perception was first raised by Wynne (1980). He argued that differences between expert and lay constructions of risk might depend upon the evaluation of the trustworthiness of risk managing institutions (Vlek and Stallen, 1981), and of the authorities to act both in the public interest and with regard to best possible technical standards and practice. One interpretation which might then be placed upon several of the qualitative dimensions of risk identified in the psychometric studies (for example, control, equity, voluntariness, known to science) is that they may tap concerns about the processes of hazard management (Bord and O’Connor, 1990).

Trust is defined in various ways in the literature, but is typically discussed in terms of an implicit relationship between two or more parties. One of its functions is to reduce complexity in our social environment (Barber, 1983), hence making our interaction with others more predictable.
In relation to risk communication, Renn and Levine (1991) list five attributes of trust:

i) competence (do you have the appropriate technical expertise?)
ii) objectivity (are your messages free from bias?)
iii) fairness (are all points of view acknowledged?)
iv) consistency (of your statements and behaviour over time?)
v) faith (a perception of your good will?)

Renn and Levine argue that trust underlies confidence, and where this is shared across a community one has credibility. Very recent empirical research in the US, concerned with the specific issue of hazardous facility siting, appears to confirm the hypothesis that trust is related to community level acceptance of hazard (Flynn, Burns, Mertz and Slovic, 1992; Slovic, 1993), although the link may not be a direct one. Bord and O’Connor (1992) argue that trust in industry is linked to trust in government and in turn to a belief that specific hazards are amenable to control. A link with perceived “control” suggests trust arguments may underlie part of this aspect of the qualitative profile of a hazard (which, as we have argued earlier, appears to hold a rather complex relationship with the other dimensions from the psychometric studies). Flynn et al (1992) find that distrust in waste repository management is positively correlated with gross perceived risk ratings and opposition to siting (see also Kunreuther, Fitzgerald and Aarts, 1993).

Clearly there is much more fundamental work to be done on this question, and in particular to establish:

a) whether the trust finding is issue specific (to hazardous waste) or generalises to other high Dread/Unknown activities in the area of health and safety;

b) whether it is as salient in the UK context as it appears to be in the US - some authors (Slovic, 1993; and Earle and Cvetkovich, 1995) discuss whether distrust is a particular manifestation of US socio-legal relationships, but Petts (1994) argues that it is indeed a key issue in the UK too;

c) how presentation of uncertainty surrounding risk assessment is mediated by questions of trust (Johnson and Slovic, 1995);

d) and the relationship between trust and perceived control over a hazardous activity, and between trust and wider discourses about responsibility, blame and security.

3.4.4 Conclusions

The broad conclusion to draw from the social and cultural approach to risk is that it has provided layers of theoretical complexity rather than direct empirical advancement. That is, we now recognise that people’s judgements of characteristics of hazards (as investigated in the psychometric approach) might be most usefully interpreted in relation to value and moral issues, involving questions of responsibility, trust and blame when things do go badly wrong. The dilemma raised by the cultural approach also remains (unsolved in Royal Society, 1992, as it was in 1983): namely, that if diverse (perhaps even contradictory) discourses or world-views do have fundamental importance for how people evaluate risk, how can these be reflected in public safety decisions other than through the rather obvious course of the democratic process? On the other hand, and as noted above, failure to find strong sociodemographic differences in risk perceptions weakens at least one objection to the adoption of group average data in expressed preference approaches such as contingent valuation. The qualitative analysis of the data from the current programme of research is intended to throw further light upon this very difficult issue.
CHAPTER 4 CROSSING DISCIPLINARY BOUNDARIES

In the previous two Sections, the economic and the psychological (and sociological) approaches were reviewed separately, and for the most part this is a fair reflection of how these two approaches developed through much of the ‘70s and ‘80s. Increasingly, however, (some) researchers in each discipline became aware of some of the questions and issues being discussed by those working in other disciplinary traditions. Often this awareness involved a degree of discomfort, with certain fairly basic beliefs and assumptions being exposed to critical scrutiny and unsettling evidence - as we shall see below. However, one encouraging result has been the recent development of lines of research which try to take greater account of a range of lessons, insights and concerns generated by the different branches of social science.

4.1 THE CHALLENGES TO THE ECONOMISTS’ APPROACH

4.1.1 Problems with Contingent Valuation

Recall from the Introduction the basic assumption underpinning standard economic theory - namely, that society is composed of individuals who know their personal preferences quite precisely, have good information about the qualities and prices of the alternatives available to them, and process this information so as to spend their money optimally, so that there is no adjustment they could make which would give them greater well-being. Thus, when asked to consider some safety improvement, such individuals are able to judge how far it would be worth their while to adjust their other consumption in order to release money to purchase the extra safety. The maximum amount that each individual would be willing to pay for the safety benefit in question provides a measure of the value of that improvement to that individual. By asking a representative cross-section of the population to state their personal WTP, and by aggregating their responses, we derive a figure for the value of preventing a “statistical” fatality (or injury, as the case may be). That, at least, is the theory.

However, it is clear from Section 3 that psychological research raises serious doubts about how far, in practice, individuals can or do process information in the way the economic model supposes. At a general level, and as noted earlier, certain eminent psychologists have offered a critique of the standard maximisation assumptions of economic theory - the most prominent of these being Simon (1955, 1957, among others). Simon’s basic argument was that optimisation was in general far too demanding a task, and that much human behaviour was better characterised by “satisficing” - trying to attain some (very probably suboptimal) “aspiration level”, and then either settling for it if it could be achieved, or else modifying it until it became attainable. This critique of conventional economic assumptions has been supplemented by the work on “heuristics”, “biases” and other confounding factors described in the previous Section.

In the particular context of contingent valuation, a number of phenomena have been identified, described in detail in the next subsections.

Oversensitivity to theoretically irrelevant monetary cues. In particular, “starting point bias” (Rowe, d’Arge and Brookshire, 1980; Boyle, Bishop and Walsh, 1985; Dubourg, Jones-Lee and Loomes, 1995) and “range effects” (Dubourg et al, 1995). The undesirable consequence of such effects is that individual responses may be systematically - and substantially - influenced by whatever stimuli the researchers choose to use. For example, Dubourg et al (1995) report a study where values for road injury risk reductions were elicited by a process of “iterative bidding”: that is, respondents were initially asked whether they would pay some amount x for a given risk reduction, and then, depending on their response, the value of x was increased or decreased until the range within which their WTP lay was established. To check for any starting point bias, half of the sample were initially asked if they would pay £25,
while the other half were initially asked if they would pay £75 Under the standard economic assumptions, this manipulation should make no significant difference; but in fact, there was a strong starting point effect - in ten comparisons, the mean WTPs from those respondents initially presented with the £75 question were all between 90% and 190% higher than the corresponding mean WTPs from those respondents initially presented with the £25 question.

*Undersensitivity to theoretically relevant information.* Recall that in the Introduction it was argued that people whose behaviour conforms with the standard model should a) be willing to pay more for any given reduction in risk as the severity of the injury increases and b) be willing to pay more for greater reductions in the risk of any particular injury. However, in the context of CV for environmental goods, Carson and Mitchell (1995) showed a "part-whole" bias in which respondents gave approximately the same WTP for a "large" good as they did for a much smaller good that was a subset of the large good. Kahneman and Knetsch (1992) called this insensitivity to magnitude the "embedding effect", for which there is now plentiful evidence in the environmental context.

Suggestive evidence of this insensitivity in the context of safety has been around for some time, in that studies which have employed smaller risk reductions have tended to obtain larger estimates of the value of statistical life (Mishan, 1985; Smith and Desvouges, 1987; Viscusi, Magat and Huber, 1987; Lin and Milton, 1993). And although Baron and Greene (1996) found that the size of the embedding effect could be reduced by reminding subjects about other goods of the same type, both Jones-Lee, Loomes and Philips (1995) and Covey, Jones-Lee, Loomes and Robinson (1995) found clear-cut evidence that WTP for one risk reduction may be no higher than WTP for a much larger risk reduction or package of risk reductions. Indeed, the relative insensitivity of CV responses to the magnitude of risk reductions is evident in the Summary Table at the end of Appendix A (see pp. 134-135) which gives the values of statistical life (VOSL) obtained from various CV studies. If log10 of each estimate is plotted against log10 of the risk reduction employed in the study concerned then the picture emerges which is illustrated in Figure 1 on the following page.

Furthermore, if log10 VOSL is regressed on log10 RISK REDUCTION then we obtain:

\[
\begin{align*}
\log_{10} \text{VOSL} & = 4.14 - 0.52 \log_{10} \text{RISK REDUCTION} \\
(0.48) & (0.10)
\end{align*}
\]

The figures in brackets are standard errors and indicate that both the constant and the coefficient of \( \log_{10} \) RISK REDUCTION are very highly significant. In interpreting these regression results, it should be appreciated that, setting aside other considerations such as international variations, complete sensitivity of CV responses to the size of risk reduction would be reflected in a coefficient of \( \log_{10} \) RISK REDUCTION that was insignificantly different from zero, while complete insensitivity would be reflected in a coefficient that differed insignificantly from minus 1. To the extent that the estimated coefficient is approximately -0.5, the results summarised in the Table at the end of Appendix A can be taken to reflect partial, though not complete, insensitivity.

The Jones-Lee, Loomes and Philips (1995) study also revealed considerable insensitivity of CV responses to differences in the severities of injuries. Prior to eliciting WTP values,
respondents were asked to rank the different injuries from least to most severe, and then to locate them on a hundred point “visual analogue” scale, using the distances between locations to indicate how much more severe one injury was than another. While there was substantial consensus about the ordering of injury states, with the distances on the scale suggesting that the great majority of respondents perceived clear differences in severity, their CV responses failed to reflect this. In one comparison between two injuries separated by more than 15 percentage points on the visual analogue scale (VAS), only 53% of respondents stated a strictly higher WTP to reduce the risk of the more severe injury by a given amount; and in a second comparison between two other injuries located 14-15 points apart on the scale, only just over 40% of the sample stated a strictly higher WTP to reduce the risk of the more severe injury by a given amount.

Figure 1  $\log_{10} \text{VOSL vs } \log_{10} \text{RISK REDUCTION}$

(Data Source - Appendix A, see pp. 134-135)
Differences between willingness to pay and willingness to accept. Although standard theory would allow that there may be some difference between what people would be willing to pay for some marginal benefit and what they would be willing to accept in compensation for the corresponding marginal disbenefit, there have been numerous papers reporting what Knetsch and Sindén (1984) referred to as an “unexpected disparity” between the two measures. In the context of health and safety, substantial disparities between the sums people state that they are willing to pay for marginal reductions in risks of death or injury and the amounts they say they would be willing to accept as compensation for tolerating marginally higher risks have been reported by Thaler (1980), Jones-Lee, Hammerton and Philips (1985), Viscusi, Magat and Huber (1987) and McDaniels (1992), among others.

There have been two types of defence for the standard economic model in the face of such evidence. At the theoretical level, Hanemann (1991) has suggested that even quite substantial disparities may be entirely consistent with standard preferences in cases where there are only very imperfect substitutes for the good being valued - and he has cited health as one possible example of such a case. At the practical level, Brookshire and Coursey (1987) and Coursey, Hovis and Schulze (1987) have suggested that the disparities are principally due to inexperience and lack of incentives to reveal true considered preferences, arguing that if elicitation procedures were designed so as to provide adequate experience, feedback and incentives, the differences between WTP and WTA responses would reduce to levels consistent with standard theory.

However, neither of these defences has been thoroughly tested, and there is as yet little clear evidence to support them. Whereas, by contrast, there has been further evidence (Dubourg, Jones-Lee and Loomes, 1994) that the disparity applies not just to people’s “point estimate” responses, but even to “personal confidence intervals”. That is, instead of being asked just for a single WTP or WTA figure, respondents were also asked in each case for upper and lower bounds, such that even if they could not be sure of their precise WTP or WTA, they felt confident that their values lay somewhere in the interval between the bounds. However, the striking result was that for a substantial proportion of respondents there was no overlap whatsoever between the intervals produced by the WTA questions and those generated by the WTP questions: that is, out of the 152 individuals who gave usable responses, the lower bound on their WTA strictly exceeded the upper bound on their WTP for 52% of the sample in the case of one injury risk reduction and for 56% of the sample in the case of another injury risk reduction.

Such disparities - not just between point estimates but between whole intervals - again suggest that something more than simple imprecision is at work. Informal debriefing of some of the participants gave some insights into what may have been happening. When asked how much they would be willing to pay for the risk reduction, a number of respondents appeared to focus on the fact that the baseline risk was already small, so that even if they took no extra precautions they felt that they were very unlikely to have an accident anyway; and thinking in this way, they were often reluctant to contemplate paying very large sums to increase survival probability by what seemed to be such a relatively small amount. By contrast, the willingness to accept question tended to shift their focus of attention: now they were more likely to think in terms of how they would feel if they did sustain an injury, knowing that there was a real possibility that the injury had been sustained as a result of accepting extra risk in return for money. From this perspective, the risk change seemed relatively large and had stronger overtones of personal culpability, and it therefore appeared quite natural to demand much greater compensation for the extra risk - or even to say that they would refuse to accept the extra risk at any price.

This kind of account of WTP-WTA disparities would seem to have much in common with the view that they are primarily due to “loss aversion” (see Kahneman, Knetsch and Thaler, 1991). The basic idea is that, rather than evaluating outcomes in absolute terms as if from some neutral standpoint, people tend to code them as gains or losses relative to some
reference point, and then treat them asymmetrically, with perceived losses looming larger than corresponding gains. This notion has quite recently been incorporated into an “alternative” model of preferences - Tversky and Kahneman’s (1991) Reference Dependent Utility model. The predictions of this model, and its implications for public policy towards health and safety, will now need to be considered alongside the accounts offered by Brookshire, Coursey, Hanemann and others, mentioned earlier.

In the paragraphs above, we have described several serious problems for the conventional CV approach which were first observed in the context of environmental questions, but which were also found to occur in questions about health and safety. Unfortunately, the health and safety context appears fraught with several further difficulties, as follows.

Communication/comprehension of small probabilities. Even in an area such as road transport, which involves higher probabilities of fatality than almost any other single cause of accidental death, the baseline probabilities for most people in the UK are quite small, and therefore the sizes of risk reduction that people can be asked to pay for are even smaller. Jones-Lee, Loomes and Philips (1995) used showcards with 100,000 small squares, of which 8 were shaded in the top left-hand corner to convey a baseline risk of death of 8 in 100,000 per year, while 4 were shaded in the top right-hand corner to convey the residual risk after a reduction of 4 in 100,000. However, it is far from clear how effectively this conveyed the risk reduction to many respondents, nor how far they were taking these data as relating to their personal levels of road transport risk.

Moreover, with a reduction of the order of 4 in 100,000, modest imprecision in people’s responses can become magnified into quite substantial differences in the corresponding value of statistical life. For example, someone who stated a “personal confidence interval” running from a minimum WTP of £100 up to a maximum of £150 would be calculated to have a VOSL of between £2.5m and £3.75m. And if (as the evidence cited earlier suggests) these intervals are insufficiently sensitive to different magnitudes of risk reduction, then smaller risk reductions are liable to produce much larger and wider ranges of VOSL. Thus if the individual in the above example were to respond to a reduction of only 1 in 100,000 with an interval running from £60 to £100, the implied VOSL would range from £6m to £10m. And when (as in the food safety study undertaken by Covey et al, 1995) the risk reduction is just 1 in 6,000,000, then the difference between a response of £5 and a response of £10 translates into a difference in VOSL’s of £30m (i.e. the difference between £30m and £60m).

The distribution of health and safety effects. It is extremely hard to design questions which can pin down different types of distributional consideration. In particular, it is hard to distinguish “safety-focused” concerns about the distribution of physical risks per se from more general concerns for the resulting distribution of welfare or utility. It may not be surprising, therefore, that no-one has yet managed to design a reliable survey method for eliciting “willingness to pay for other people’s safety”. There have been some interesting preliminary studies, however, which have examined three main issues.

1: Statistical vs. anonymous death. “Statistical death” refers to a distribution involving statistically independent risks of death to each individual; “anonymous death” refers to a distribution involving certain but ex ante anonymous death. Hammerton, Jones-Lee, and Abbot (1982) found that most respondents preferred “anonymous death”; Loomes (1982), however, found considerable variation in distributional preferences and, indeed, considerable variation in the qualitative reasons given for these preferences.

2: Ex ante vs. ex post equity. “Ex ante” equity refers to equality in the prior probability of harm; “ex post” equity refers to equality in the final outcome. Keller and Sarin (1988) found that respondents were concerned with both notions of equity, which featured with different degrees of prominence in different contextual scenarios.
Risks vs benefits. Keller and Sarin (1988) found that preferences for risk distributions depend also upon the distribution of benefits. Keller and Sarin (1995) found that most respondents placed considerable emphasis on balancing the distribution of risks and benefits irrespective of cost-effectiveness considerations.

The timing of health and safety effects. Discounting health benefits for time inevitably raises the thorny philosophical issue of how to deal with benefits to future generations. There are two distinct traditions in the economics literature about this matter. According to the dominant tradition in the UK, the discounting of benefits to future generations should represent an ethical judgement on behalf of all generations from the point of view of an "impartial observer" (Broome, 1992; Parsonage and Neuberger, 1992; Jones-Lee and Loomes, 1995a). According to the dominant tradition in the US, however, the discounting of benefits should represent an ethical judgement made on behalf of the current generation only. It is perhaps not surprising, therefore, that the only research which has used questionnaire methods to elicit the time preferences of the current generation for physical safety benefits has come from the US (Horowitz and Carson, 1990; Cropper, Aydede and Portney, 1992).

This US research has used pairwise choice questions about public policy scenarios involving different numbers of deaths at different times. The central finding is that, contrary to standard economic theory, individuals do not appear to discount health effects at a constant rate. The health discount rate appears to be extremely high for short time horizons, low for longer time horizons, and to vary greatly between individuals and between situations. These results are reinforced by other work by psychologists on intertemporal decision-making, which has focused on private consumption choices rather than public policy choices - see, for example, the collection of papers in Loewenstein and Elster (1992).

The difficulty/unacceptability of trading off money and health. Although people may be able to handle trade-offs between goods that are reasonably similar and have a number of dimensions in common (e.g. the characteristics of one form of entertainment compared with another), such trade-offs become more difficult as the comparators become more dissimilar. Thus a direct comparison between money (or the other things money is being spent on) and unfamiliar small probabilities of unfamiliar states of health is a very difficult and demanding task.

Moreover, in some contexts, a significant number of individuals may be unwilling to consider a direct money-health trade-off for ethical or cultural reasons. For example, when Propper (1988) asked people about their willingness to pay for shorter waiting times for non-urgent surgery, she encountered numerous objections from people who interpreted the question in terms of "jumping the NHS queue", and who either refused to answer, or else recorded "protest" responses of zero. In a society where people have come to expect public provision of health care "according to need", funded from general taxation rather than paid for at the point of consumption, researchers have often encountered resistance to CV questions for particular health care interventions. One response to this has been to explore other non-CV techniques for eliciting preference indices. In the next subsection, we briefly review several of these, and consider some of the pertinent issues raised.

4.1.2 Non-monetary Measures of "Value"

In the health care field in particular, a number of techniques have been developed which seek to establish "Quality of Life" (QoL) indices to measure the relative (un)desirability of different states of health on the spectrum from "best possible state" to "death", and even beyond, to "states worse than death". In some applications, such QoL indices have then been combined with the periods of time spent in various states in order to calculate "Quality Adjusted Life Years" (QALYs) - that is, to convert any particular health profile into the
“common currency” of years of good health. On this basis, the benefit of any intervention may be quantified in terms of how many QALYs it adds, and alternative interventions may then be ranked on the basis of “cost-per-QALY”.

Of course, in order to put a monetary value on benefits measured in this way, it would be necessary to find some method of valuing a QALY in monetary terms. This is an issue to which we return later. But first, we focus attention on three techniques often used to estimate QoL indices, and the issues arising from this work.

**Visual Analogue Scaling (VAS)**
A visual analogue (or rating) scale often consists of a single line (perhaps calibrated like a thermometer) on a page, and requires the respondent to locate their most preferred health state at one end of the line and their least preferred state at the other. The respondent then rates the desirability of each intermediate health state by placing it at some point on the line between these two “anchors”. If such scales are to yield interval measures then the relative distances between pairs of health states must reflect how the respondent perceives the difference between one pair of health states as compared to the difference between another pair. Thus if a state B is placed midway between two states A and C, this is supposed to reflect the fact that the respondent regards being in state A as better than being in state B to the same extent as being in state B is better than being in state C. If the position occupied by “normal health” is assigned a rating of 1 and the position occupied by death is assigned a rating of 0, indices can be computed for all other health states according to their locations on the line. States rated as worse than death are assigned negative indices.

**Time Trade-Off (TTO)**
The procedure here is to present a respondent with some given period of time - 10 years, say - in a particular health state (followed by death at the end of that period) and ask how much of the 10-year period the respondent would give up in order to achieve normal health for a shorter period of time (also followed by death). If the respondent states that s/he would give up the last eighteen months of the 10 years in health state A in order to have normal health from now until death in 8.5 years time, health state A is accorded a QoL index number of 0.85.

**Standard Gamble (SG)**
The simplest form of SG question presents a respondent with a choice between the certainty of being in health state A for the rest of their life, and an alternative “treatment” which, if successful, will improve their health state (in the simplest case, restoring them to normal health for the rest of their life) but which, if it fails, will put them in a worse state (in the simplest case, immediate death). The respondent is then asked to identify the chances of success (and the corresponding chances of failure) that would make him/her feel that the two alternatives are equally preferable. Thus if this “point of indifference” between the two alternatives is reached when there is a 95% chance of success (and hence a 5% chance of failure), health state A is assigned an index number of 0.95.

There are a number of other techniques besides the three outlined above, and more comprehensive surveys can be found in Froberg and Kane (1989) and Brooks (1991). However, the three above have been the most widely used and discussed, and the body of research using one or more of them has raised several crucial concerns.

First, the three techniques tend to produce systematically different QoL indices for the same health states. In particular, VAS often assigns markedly lower indices, especially to less severe health states, while SG often assigns the highest indices, with TTO somewhere between, although typically closer to SG. In one respect, this may not be surprising, since there is no theoretical reason why they should generate the same numbers; as discussed in Loomes (1994) and Robinson, Jones-Lee and Loomes (1994), VAS and SG methods have
distinctive theoretical foundations, while TTO indices at least, when derived in the simplified manner outlined above - require very restrictive assumptions if they are to be given a sound theoretical justification.

However, as the introductory review in Loomes (1994) also shows, even the method which is supported by the conventional theoretical framework - namely, SG - has also been found to be not entirely immune to the possibility of generating internal inconsistencies, under certain sets of circumstances. And subsequent work (Robinson, Jones-Lee and Loomes, 1994; Robinson, Dolan and Williams, 1997) suggests that the internal consistency and robustness of VAS and TTO indices may also be questionable. Thus, there are at least some doubts as to the ability of these procedures, or, indeed, any of the other techniques discussed in Brooks (1991) to provide a wholly secure and general basis for estimating even the relativities between health states.

Despite this, there have been some attempts to harness QoL indices in order to compute a monetary value on loss of health status due to illness or injury. What has been called the Relative Utility Loss Approach (RULA) proceeds broadly as follows.

First, typical health profiles resulting from illness or injury are identified by relevant experts. Then some set of QoL indices is used to calculate the utility loss entailed in any profile as the difference between the index number for normal health, 1.0, and the QoL index for the state(s) that make up the profile, weighted by the length of time spent in each state worse than normal health (perhaps with some discount factor reducing the weight of periods of time further in the future). This is, in effect, a calculation of QALYs lost.

The next step is to derive a monetary value for a QALY. In two recent studies - Ives, Soby, Ball and Kemp (1993) and Davies and Teasdale (1994) - this has been done by taking the pure willingness to pay component of the figure used by the Department of Transport to value the prevention of a statistical road fatality, estimating how many lost QALYs such a fatality would entail, and computing the annuity which (under various assumptions) would reconcile the two. By this method, Davies and Teasdale arrive at a figure of £27,150 per QALY lost (1990 prices), which they then combine with the RULA procedure to generate values for preventing various work-related accidents and causes of ill-health. Ives et al did something similar in principle to estimate values for the prevention of non-fatal road accidents, as did Miller, Calhoun and Arthur (1990), while Mauskopf and French (1991) used this approach in the area of food safety.

At first sight, this approach appears attractive. It is relatively cheap, and can value a wide variety of different kinds of health and safety benefit using the results of existing valuation studies. Against that, the results of the method are sensitive to the choice of health state index and to the computational assumptions. The studies by Mauskopf and French (1991) and Ives et al (1993) both found inconsistencies between the results of using different sets of “off-the-peg” health state indices. Moreover, even when some of the sets of indices used in the Ives et al study produced reasonably similar results, it turned out that they were also not dissimilar to the values produced by applying CV to the same non-fatal injury descriptions, which Jones-Lee, Loomes and Philips (1995) showed to be subject to a number of serious upward biases, and which the Department of Transport therefore rejected as a basis for valuing the benefits of preventing non-fatal road injuries, preferring instead to use estimates derived from custom-built SG questions and linked directly to the value for preventing a fatality.

Finally, another reason for being wary of the RULA approach relates to the question of transferability, mentioned in the Introduction. The estimation of health state indices is often done under conditions either where there is no strong sense of context or else where the context is medical/surgical - i.e. (health care interventions for) existing health states whose origins may not be well specified. Given the extent to which the literature reviewed in Section 2 suggests that risks of death on the roads may be regarded very differently from risks of
death from other hazards, it is doubtful how far the value of preventing a road fatality can be expected to apply in other contexts. It is even less clear that cross-breeding the road-based VOSL with some acontextual or medical/surgical-based set of health state indices will provide a sound basis for valuing the prevention of non-fatal injuries or ill-health in other contexts. A warning note along these lines is sounded in paragraph 25 of Appendix 6 in the Department of Health’s 1995 document Policy Appraisal and Health.

One other non-monetary approach to eliciting safety relativities is the Risk-Risk or Risk Trade-Off (RTO) method. Within a particular context, RTO questions may require respondents to trade off a small change in the probability of one severity of injury or illness against a small change in the probability of an injury or illness of different severity. Across contexts, RTO questions may require respondents to balance the risk of death from one hazard (air pollution, perhaps) against the risk of death from another hazard (such as road accidents).

The advantage of RTO questions is that they focus on the appropriate probabilities so far as standard economic theory and the realities of health and safety control are concerned - namely, marginal variations in already small probabilities. But herein lies the main practical drawback: if it is difficult to communicate/comprehend a marginal reduction in a single small probability, it is potentially even more difficult for respondents to balance two marginal changes in two small probabilities, especially if the exercise involves a marginal increase from one low baseline being offset by a marginal decrease from another (different) low baseline.

Two studies using RTO (Dolan et al., 1995, and Covey et al., 1995) gave particular grounds for concern about the robustness of relativities produced by this method: not only was there evidence of serious inconsistencies in the recorded responses, but “debriefing” discussions with respondents also suggested that they had considerable difficulties in dealing with the task, and were liable to resort to some inappropriate heuristic. Against that, Viscusi, Magat and Huber (1991) and Krupnick and Cropper (1992) report no direct expressions (or impressions) of widespread unease or misunderstandings. Nevertheless, they found it necessary to trim about one third of the responses from the sample on the grounds of inconsistency. Furthermore, they found it necessary to use median ratios, rather than the theoretically appropriate mean ratios, to arrive at what they considered to be plausible and consistent results.

Overall, then, it appears that there is a substantial body of evidence showing that in a variety of forms the standard assumptions of the economic model either do not hold or else are not easy to operationalise in applied research. While such evidence may be disconcerting for those rooted in the mainstream of the economics discipline, much of it would come as no surprise to psychologists: indeed, as we have seen, it has often been psychologists who have demonstrated the vulnerability of the conventional economic model to a variety of biases and inconsistencies. However, to date psychology has not been able to offer a substitute for the economic model which can provide an alternative basis for the allocation of scarce resources to various health and safety policies. In the next subsection, we therefore consider the psychological approach - both its insights and its limitations - in more detail.

4.2 THE CHALLENGES TO THE PSYCHOLOGISTS’ APPROACH

Probably the single most influential programme of risk research emerging from the psychological literature has been the psychometric work on the “qualities” of hazards, initiated by the Oregon group and replicated and elaborated by many other researchers during the past fifteen years. The well-known Dread/Knowledge diagram appears to organise the evidence in a plausible and parsimonious way, in that as one moves from bottom left to top
right, hazards are perceived as progressively more risky, less tolerable, and more in need of regulation and control.

However, as a basis for guiding the allocation of scarce health and safety resources, this diagram has important limitations. It might be tempting to suppose that one could simply superimpose some form of expenditure contour map on the Dread/Knowledge diagram, such that the VOSL gets progressively higher as one moves from bottom left to top right. If this were possible, it would allow policymakers to simply read off the appropriate VOSL for any particular hazard.

Unfortunately, there are a number of reasons for doubting whether such an appealingly simple approach would be valid and/or reliable.

One potential difficulty arises from what was referred to in Section 2 as the “Russian Doll” effect, whereby any one generic hazard located at a particular point on the Dread/Knowledge map may consist of a number of more specific hazards which are quite broadly distributed across all four quadrants of the factor space. This may not be an insuperable difficulty, but at the very least would require policymakers to decide upon the level of aggregation to be used in valuing benefits: for example, whether to use a single value for the prevention of any and all road accident deaths, or whether to distinguish between drivers and pedestrians, children and adults, those who have been drinking and those who are the victims of drunken drivers, and so on.

A second, and perhaps more fundamental, problem is that although the factors subsumed in the Dread/Knowledge representation may be important in driving perceptions, there may not be a direct mapping from perceptions to preferences/values. Put another way: how far do the factors that enter into perceptions also reflect values that should influence public expenditure and/or regulation? This is a large question, which we now consider in some detail.

To allow the perceptions of the public to influence policy has seemed to some to be a great mistake. It has been argued that members of the public tend to be ill-informed (and even irrational), and thus to input their perceptions into policy would be, at best, to introduce noise into the system, and, at worst, to introduce systematic bias. Lives will be lost, so the argument goes, as volatile public preferences and perceptions lead to scarce resources lurching between “hazards of the month”.

To state our own position from the outset, we are far more optimistic than this; we believe both that the public’s input is in principle important for societal decision making (Royal Society, 1992; Pidgeon, 1996a; Pidgeon and Beattie, 1998; see also National Research Council, 1996), while aspects of perceptions can also be used to greatly enrich expert risk analyses. However, we first briefly review the arguments against this position. The remainder of the first part of this section will then concentrate on elucidating a set of conditions that will be most likely to maximize the useful contribution of public perceptions, while minimizing their potentially damaging aspects.

Arguments against allowing the public’s risk perceptions to influence public expenditure are outlined in the following subsections

4.2.1 “Risk Perceptions are Noise/Bias”

Probably the most frequently argued thesis is that perceptions have no place in policy decisions because the public does not have the knowledge to accurately perceive risk, and thus judgements will be noisy or biased. As a result, systematic error will be introduced into the (appropriate) risk analyses performed by experts (Okrent, 1996).
Certainly, people's perceptions can be erroneous, transient and biased. Fears of the terrorist risk in London taken one day versus one week after an event such as the IRA bomb attack on Canary Wharf would clearly show their volatile nature. However, such effects are entirely predictable from the psychological theory of availability (Kahneman, Slovic and Tversky, 1982), and thus may be controlled for by appropriate timing of value elicitation. Perceptions can be also erroneous due to lack of information - it would be unreasonable to expect laypeople to deliver informed judgements off the cuff concerning the risks from accidental releases of genetically modified organisms (GMOs), for example. However, this does not imply that all public values are ill-informed. For example, people may be in possession of much relevant information concerning their own individual risk of domestic fire. This criticism might also be overcome by participative education, such as discussion groups in which people are given relevant information by experts (as, for example, citizen panels - Morgan et al., 1994). Psychological theory would also predict that discussion groups would have other benefits, such as reducing framing effects by encouraging alternative framings of the questions, improving reflection through public justification of arguments, and providing additional information. Thus, while much methodological work remains to be done, the case that public perceptions are inherently too error-prone to be of use is certainly not proven.

4.2.2 “Perceptions Might be Prejudiced/Discriminatory”

A potentially serious charge that can be levelled against risk perceptions is that they are likely to reflect prejudices against particular social groups. This can lead to public policy instantiating undesirable, or even illegal, allocative procedures. While there is a plethora of psychological literature showing negative stereotyping of out-groups (for a review, see Sabini, 1995), there has been little linking this to allocation decisions. A recent article by Irwin, Jones and Mando (1996) extended Slovic et al.'s (1988) psychometric approach to look at the dimensions along which victims of risk (in this case of HIV infection) are represented. Perceived distastefulness and riskiness of the method of infection and the "deservingsness" of the victim largely explained the representation of the victims. Irwin et al. point out that these results can be related back to some of Slovic et al.'s findings: for example, that illicit drugs promote less concern than pesticide risks. While this has been explained in terms of the voluntariness dimension, it is also possible that the likely victims (e.g., drug abusers vs general public) promote differential degrees of concern, and hence would be differently valued by society.

To date little research has established the degree to which perceived differences in victim desirability are likely to influence public values, but it would be surprising if this were not the case. How to deal with this issue is an unsolved problem. In some cases such differences may seem to be "legitimate" bearers of wider social value (e.g., valuing children differently from old people). Qualitative data from group discussions may yield insights into such processes, although many unsavoury prejudices will be held unconsciously, or will not be voiced in public.

4.2.3 “Managing Risk Merely Entails Managing Perceptions”

If perceptions were the only input into public sector allocative decisions, the public might be open to abuse by unscrupulous risk managers. Rather than regulating and reducing risks, risk managers might concentrate efforts (and resources) instead on trying to convince the public that the threat is negligible (see Rimmington, 1995).

Certainly the public have been manipulated in the past by self-interested parties, as when companies have sought to cover up public hazards. However, few have claimed that public perceptions should be the only input into allocative decisions - expert judgment and regulation of risks must also play a part.
A corollary of the position argued in (3) is that undue attention to perceptions may lead to inequitable exposures: sectors of the public who perceive the risk as being lower might be asked to bear more risk than other sectors. Again, governments can intervene to prevent this, if it is deemed undesirable (which is of course a political or ethical judgment).

The three arguments above present the main threats to the position that allocation of resources should reflect public perceptions of risk. We believe that none is so damning a criticism as to suggest that perceptions should be largely discounted. Instead, we would conclude that steps should be taken to ensure that the public from whom judgments are elicited should be in groups, with access to properly explained technical and scientific information, and that efforts should be made to detect prejudiced perceptions with a view to later debiasing. Care should also be taken to reflect the diversity of viewpoints in our culture. With those considerations in mind, we now turn to the arguments in favour of taking account of public risk perceptions, from which we shall draw more lessons concerning the “best” conditions in which to elicit these perceptions.

We can identify four arguments in favour of allowing risk perceptions to influence public expenditure and safety regulation, which are outlined in the following subsections.

4.2.4 “Perceptions have Consequences”

This argument suggests that policy ignores public perceptions at its peril, because these perceptions will lead to actions (however rational or otherwise such actions may be thought to be by the experts), and these actions will have consequences. For example, risk analyses may determine that public licensing of youth activity centres is unwarranted, yet public outrage (following the Lyme Bay tragedy - see House of Commons, January 27th, 1995) may lead the public to withdraw its custom unless regulation is imposed. Failure to license thus might cause activity centres to go out of business (or politicians to fail to be re-elected). Similarly, risk analyses might determine that it is unnecessary to evacuate civilians following a particular low-level radiation leak at a nuclear power plant. Yet the news of such a leak could produce widespread panic and flight, causing more harm than the leak itself would have been expected to produce. Thus in hindsight, evacuation (however “irrational!”) might have been expected to produce the best outcome. This argument indicates that, at minimum, policymakers must show awareness of the public’s likely perception of a given risk, and be prepared to act on the perception if the risk produced by the perception will exceed that from not acting. Indeed, it would be quite possible to take account of such factors in a conventional risk assessment. However, to do so requires both a good understanding of their operation, and a willingness to build them into the analysis.

4.2.5 “The Public Should have Input into Risk Decisions that Affect the Public”

Sir Frank Layfield, in his report following the Sizewell-B inquiry, stated that ‘The opinions of the public should underlie the evaluation of risk’ (Layfield, 1987, summary paragraph 2.101h). This has been instantiated in HSE’s philosophy on the tolerability of risk (ToR) (HSE, 1992), which states that the public should not be asked to live with a risk that is viewed as intolerable (defined in ToR as a risk of fatality of greater than 1 in 10^-4 per annum for the public, and 1 in 10^-3 per annum for workers). It is also recognised that, for certain risks (e.g. nuclear power - see Rimmington, 1995) the public’s tolerability limit, and with this the criterion set by the regulator, may be found to be much lower. Public input is also currently achieved more directly when expressed preference studies are used, as discussed at length in the current document, as the foundation for VOSL criteria used in public policy and other decision making processes. The psychometric work on risk perception clearly sheds light on this issue.

Beyond setting tolerability limits, there are other ways in which risk perceptions may be taken into account, and other reasons why this would be desirable. One motivation is
ethical/philosophical: exposing an individual to risk without consultation may be viewed as contrary to the democratic process or an infringement of individual rights. Such philosophical niceties may also have tangible consequences. First, broad participation in decision making (see Royal Society, 1992) increases an individual's commitment to the course of action selected. For example, even if an individual dislikes the action chosen, he or she may feel committed to it if the decision process is seen to be fair (see Rayner and Cantor, 1987). Second, participatory decision making may also increase trust in the organizations who manage the risk (National Research Council, 1989), and trust tends to increase community acceptance of hazards (e.g. Slovic, 1993).

4.2.6 “Experts can be Biased Too”

Much of the risk literature in the 1980s drew a sharp distinction between assessment of objective (correct) probabilities by knowledgeable, impartial experts, on the one hand, and subjective (possibly seriously biased) probability estimates by laypeople on the other (see Royal Society, 1992, for a discussion). It is now generally accepted (see Section 2) that such a distinction is untenable in its strong form (Watson, 1981; Fischhoff, 1989). In part this is because all risk assessment procedures require some human input from the experts performing the analysis: for example, requiring the analyst to decide which attributes should be included in the analysis, or the utility associated with particular outcomes (Pidgeon, 1996b).

Furthermore, there is now a considerable literature suggesting that it is not only the general public who exhibit systematic biases: experts may do so as well. Experts' overconfidence in their own judgments in real-world settings has been demonstrated repeatedly (e.g. Hynes and Vanmarcke, 1976; Henrion and Fischhoff, 1986; Svenson, 1989), although experts have also shown to be underconfident in other areas (e.g., Kahneman, Slovic and Tversky, 1982). Experts are also subject to availability and other biases (see Section 2). It is tempting to believe that placing experts in groups would eliminate biases, but the literature suggests that this is not always the case. Janis’ (1982) influential work on Groupthink shows through historical analysis that like-minded groups of experts can produce extreme judgments accompanied by extreme confidence, which can lead to disastrous choices (see Section 2).

Thus it seems safe to conclude that much risk evaluation is a complex decision process involving not only accepted facts, but also contested facts and individual values (Royal Society, 1992). While expertise is crucial in providing relevant information concerning both the facts and their associated degree of uncertainty, experts cannot be viewed as unbiased gold standards of judgment. The psychological literature provides prescriptions concerning debiasing of experts either individually (provide detailed feedback on the quality of an expert's predictions - see Yates, 1990), and in groups (groups should be constructed from individuals with diverging opinions in a culture of open discussion - Janis, 1982). However, expertise alone is not sufficient to ensure unbiased, representative coverage of the issues.

4.2.7 “Public Risk Perceptions could Enrich ‘Expert’ Risk Analyses”

The above analysis suggests that experts are in a privileged position in terms of information, but that their values and decision processes may not always be employed optimally. In addition, while novices may not possess all the relevant factual information, they may nevertheless be able to augment expert risk analyses with additional useful information, resulting in an overall superior analysis. There are many forms that this proposed enriching might take.

Expert risk calculations often use “expected number of fatalities” as the outcome variable of interest. The public may care about other outcome measures, such as non-fatal injuries, or “societal risk”, and hence a richer representation of the problem may be arrived at by inclusion of such values in the decision tree (see von Winterfeldt and Edwards, 1984).
Societal risk, in particular, has been defined by HSE as:

‘the risk of widespread or large scale detriment from the realization of a defined hazard the implication being that the consequence would be on such a scale as to provoke a socio/political response...’ (HSE, 1995b, p25)

However, as we argue below (Section 4.3), the issue of accident size (or Scale) may be less important in defining what is important about aversion to societal risk, than are the other qualitative psychometric factors attendant upon many large scale societal risks.

The public may also represent a plurality of values unlikely to be found in expert risk assessment. As noted above, the cultural view argues against a unitary public value system. Cotgrove (1982) makes the interesting point that risk analysts tend to be of similar cultural views (cornucopians), at least in their work lives. This suggests that other cultural viewpoints (catastrophists) will be underrepresented (and this sort of argument may be valid even if Cotgrove’s particular categorisation of value systems does not hold up empirically). In particular, the public has shown itself to be concerned with such issues as the trustworthiness of the risk managers, the feasibility of evacuation plans, and the plausibility of the probability estimates: variables which are very difficult to handle in formal risk assessments. Of course the problem remains of how multiple value systems should be integrated into an overall decision or valuation, although the democratic process or the use of decision analytic techniques both provide different sorts of solutions. It is arguable that, at minimum, public concerns should constitute qualifications to the formal use of risk assessment.

The public may also possess knowledge not readily available to experts (Wynne, 1992), such as local knowledge not available to large corporations or government. Special Interest Groups (SIGs) may have collected large amounts of data and expertise concerning the problem under investigation. While data appear to reside in the realm of objective facts, SIGs are likely to have different value systems from the establishment, which will cause their data collection to have different emphases and to place different interpretations on the data (e.g., suspected, but not formally documented, safety violations). They may also have access to different forms of information (e.g. whistle-blowers may be more likely to contact SIGs than establishment bodies).

The arguments reviewed above suggest that public perceptions may have a role to play in the valuation and regulation of health and safety. Certainly public judgments should not be the only input to such decisions, and there are clearly cases where they are likely to be error-prone or biased. However, psychological theory provides good tools for predicting such cases, and also some tools for debiasing the judgments. Debiasing can take a wide variety of forms, including group discussions, public education, varying the elicitation techniques, and decision structuring. Such methods hold no guarantee of a perfect result, but the evidence suggests that a more participatory decision process, involving both experts and members of the public, offers a richer and more complete understanding of the problem than will be offered by conventional risk analysis on its own.

However, that is a statement of general principle. The question that has not been answered so far, and which neither economics nor psychology on their own appear to be able to answer, is how we can move from a recognition that it is not simply final outcomes and their respective probabilities which drive public preferences, to developing a procedure for giving due weight to whatever other considerations are legitimate carriers of value. The next subsection reviews a number of attempts to address that question.
4.3 BRIDGING THE GAP: VALUING THE "QUALITIES" OF HAZARDS

In an early article, Fischhoff, Watson and Hope (1984) argued that the qualitative risk dimensions might provide input to risk ranking policy decisions via multiattribute utility analysis (with Dread and Knowledge as additional attributes alongside expected fatalities, injuries, physical damage and benefits). More recently, attempts which draw upon both psychology and economic theory have been made to identify whether a “context premium” is associated with people’s expressed preferences for safety improvements as a function of relevant differences along qualitative risk dimensions (voluntary versus involuntary etc.), in addition to simple assessed risk.

While the combined evidence is suggestive, it must be noted from the outset that only a small number of WTP studies directly address this issue, and that none, for different methodological reasons, are fully conclusive when taken in isolation. Although the group of studies reviewed here involve safety valuation, they may be seen as complementary to a wider research effort to critique and develop appropriate methodologies for valuing non-monetary environmental goods (Fischhoff and Furby, 1988; Fischhoff, in press).

From the earlier discussion of the impact of the psychometric factors upon risk perception and acceptance, one would first expect WTP to be highly case sensitive (as found with Gardner and Gould, 1989). That is, it will be dependent upon the particular regulatory sector (transportation, say), the hazardous activity addressed within this sector and its perceived characteristics (e.g. scheduled air travel), and local historical and contextual issues associated with that activity (recent prominent accidents, who benefits from the activity, who is exposed etc.). For example, Mendeloff and Kaplan (1989) found up to approximately twofold differences in willingness to pay to prevent a number of deaths by each of several different causes, such as workplace exposure to cancer inducing chemicals with delayed effects, bicycle and automobile accidents, and fatal crib-slat accidents to children.

More general predictions based upon the psychometric findings can be advanced as follows. First, we should expect a general trend of higher WTP for safety with hazards which evoke particular Unease or Dread (or evoke a number of significant characteristics which load onto one or both of these dimensions). Three recent studies suggest that this may be so: John, Wertenbroch and von Winterfeldt (1990), McDaniel, Kamlet and Fischer (1992) and Savage (1993). The latter author reports that this became particularly pronounced for cancer (and he reports a similar effect, in a previous study, for nuclear power generation).

Regarding the Knowledge dimension, one might initially expect Unknown risks to generate higher WTP. However, and as noted above, the relationship between Knowledge and risk evaluations is a complex one. For example, McDaniel et al (1992) report that for less well-defined (i.e. Unknown, Unfamiliar) hazards, higher WTP was related to Dread and Severity, but for well-defined familiar hazards, WTP was more closely related to Personal Exposure to the risk. Savage (1993) finds (but with one striking exception - that of stomach cancer) that WTP generally decreases where respondents believe that there is less scientific or personal knowledge about a hazard. Savage argues that his respondents might reasonably have concluded that risk reduction expenditure would have more chance of efficacy where the risk mechanisms are already well understood, and Jones-Lee and Loomes (1995b) find some evidence of similar reasoning in the qualitative responses to their recent study of underground rail versus road transport safety in London. Clearly, some fundamental research is required here regarding how understandings of the level of current knowledge of a hazard might modify WTP judgements.

In terms of more specific regulatory sectors, in the UK Brown and Green (1981) provided an early and thoughtful discussion of the psychometric characteristics and WTP. In their study, they compare WTP judgements for road and fire hazards, finding some evidence of a context premium for the latter (although their result may be confounded by anchoring response...
biases). HSE/NRPB (1989) report a pilot project looking at WTP for a range of hazards (workplace, medical, major industrial hazards) although this study suffers from some of the methodological problems discussed elsewhere in this report. Accordingly, there is difficulty in interpreting the results in isolation from response biases induced by questionnaire format (for example WTP versus WTA), and the problem of instability in final value of statistical life figures when considering very low probability risks. Gregory and Lichtenstein (1994) report higher WTP values when a general description of a possible Unknown, long-term consequence is added to two otherwise familiar hazard scenarios (cycling and automobile risks respectively). Finally, the study of Jones-Lee and Loomes (1995b) points to a general WTP context premium (of some 50%) for Underground rail over road transportation, largely attributable, they believe, to the greater sense of involuntaryness people associate with Underground journeys, and the feeling of having less control and handing over Responsibility to others when travelling by Underground.

While the evidence about the impact of the qualitative characteristics of hazards upon WTP is suggestive, that about accident size (or Scale) appears initially less straightforward, and highly dependent upon the question being asked. While HSE/NRPB (1989) do report some aversion to large-scale risks in their sample, Slovic, Lichtenstein and Fischhoff (1984) find that, when asked directly, respondents were not willing to forfeit extra individual lives in small-scale accidents in order to prevent a single large-scale accident (a result also reported by Jones-Lee and Loomes, 1995b). This is not a surprising result, given the availability of strong social discourses about equality of life. Slovic, Lichtenstein and Fischhoff (1984) argue that what might be more important in producing aversion to catastrophic events may be characteristics other than size which give an event “signal value”. This is why question wording is so critical, since we would expect that aversion to Scale might not be expressed when the question is put in terms of a decontextualised or hypothetical event, but may be manifested when concrete events and their qualitative characteristics are made explicit.

In regulatory terms, if Scale is indeed less important per se, but covaries with other qualitative dimensions (Dread/Control and aspects of Responsibility/Trust) that are important in forming safety preferences, it may still be of considerable practical importance. In particular, for all practical purposes (in the absence of more definitive data concerning the influence of the other dimensions, and aside from the extra political/societal costs that large-scale events are held to generate) accident size might still be a sensible regulatory rule of thumb to adopt, because it is a heuristic for identifying where a "context" effect for stricter regulation might indeed be operating in expressed preferences. In philosophical terms, however, it is important to make a clear distinction between the two questions of context and scale.

Although the various studies discussed above have tried to go further than most “standard” CV exercises in terms of the factors they explored as possible contributors to expressed values, the majority of them were still fairly conventional in the sense that they relied heavily on eliciting answers to predetermined questions in survey format. That is, to some (often considerable) extent, they assumed that respondents comprehended the questions in the way the researchers intended and were able to produce, at relatively short notice and with only limited opportunities for deliberation and reflection, expressions of reasonably highly articulated values.

However, some writers have raised doubts about the extent to which such questions can tap into readily available “true” values and preferences: indeed, doubts have been raised about whether such well-formed and well-behaved preferences for unfamiliar health and safety (or environmental) benefits actually exist (see Fischhoff, 1991; in press).

Rather, the psychological evidence points to the fact that much reasoning with hypothetical CV questions (and other scenarios of this type) can be characterised as “constructive” (see Schkade and Payne, 1994). That is, respondents proceed by building an answer (possibly
drawing upon a small set of basic values) using a variety of context specific judgement strategies and heuristics, and in response to a wide range of cues of varying dependability. The latter would explain the sensitivities of CV responses to theoretically irrelevant features of question format, as well as insensitivities to factors that may be implicit/opaque rather than explicit/transparent in CV surveys. Under such circumstances, it is not at all surprising (indeed, it is very much to be expected) that such questionnaire-based methods will generate values that are inconsistent with one another from the standpoint of standard economic theory (Gregory, Lichtenstein and Slovic, 1993; Kahneman and Ritov, 1994).

Therefore, it has been suggested that if expressed preferences are indeed to provide a direct and reliable input into regulation and/or public expenditure policy, more intensive value elicitation methods may need to be developed. This may involve deliberately harnessing the constructive process through the use of small group discussions and intensive interactions between investigator and respondent. In this way, it might be possible to tap into basic values, upon which respondents can build more highly articulated and carefully considered responses which will better represent their interests. Proposals about how to achieve this have recently been made (see e.g. Gregory, Lichtenstein and Slovic, 1993; Morgan, Fischhoff, Lave and Fischbeck, 1994; also Weble, Rakel, Renn and Johnson, 1995) although no full-scale study has yet been conducted either here or in the US. Nevertheless, a brief review of the literature relevant to the “constructive preference” approach might be an appropriate way to conclude this review.

4.4 CONSTRUCTIVE APPROACHES TO VALUATION

Three variants of a more constructive type of approach are “value-focused thinking”, “multi-attribute utility theory” and “decision analysis”. All stress the need to decompose a decision into its component parts, and to consider the salient factors which drive decisions or (as in our case) carry value. Such an approach has the advantage of providing checks for the internal consistency of responses and emphasizes the dimensions of the problem. In principle, monetary values can be obtained by including various levels of cost as one of the dimensions.

Decision analysis can be used to attach values to alternative policy options using explicit or implicit weights and utilities (Edwards and von Winterfeldt, 1987; Keeney, 1988). Alternatively, it can be used to simply structure and inform the decision process (Baron, 1997).

Most applications using explicit values are to be found in the context of specific policy options - for instance, the siting of a nuclear power plant (Keeney, 1980a), or environmental impact assessment (Gregory, Keeney and von Winterfeldt, 1992). The key factors and weights are derived from so called stakeholders - policy makers and consumer interest groups - who are supposed to possess relatively well-formed preferences. However, it has been argued that the general public may also be regarded as stakeholders, whose views should be sought at an early stage, since delayed or peripheral involvement may lead to confrontation and collapse of the process (see e.g. Jenni, Merkhofer and Williams, 1995).

One approach to incorporating the views of the public has been through the use of citizen panels, with experts and consumer groups suggesting likely scenarios and policy options. Most of the concerns that may underlie safety decisions have already been identified in the literature, and (if we are confident in respondents' ability to articulate their values) it is then mainly a matter of getting respondents to express weights. To facilitate this, most approaches in this area attempt to keep the procedures as simple as possible, often using rankings to determine the weights (Renn, Weble, Rakel, Diezel and Johnson, 1993; Weble, Rakel, Renn and Johnson, 1995).
Breaking down a decision into simpler elements has been shown to increase consensus (Jako and Murphy, 1990). It is claimed that an important advantage of this approach, particularly in contentious areas, is its ability to reduce conflict and lead to eventual compromise. Trust and cooperation can be developed by reaching agreement on lower priority concerns, suggesting pathways for resolution of more entrenched disputes (Humphreys, 1981). Indeed, it has been suggested that most disagreements tend to be related to the implications of each policy, rather than the underlying weights and values (see Gregory and Keeney, 1994). Discussions, therefore, may serve to improve respondents' understanding of the implications for each alternative/policy. For example, Kemp and Willetts (1995) found evidence that decision analysis changed respondents' views about less well-known areas of New Zealand government policy, such as the National Library. Gregory, Slovic and Flynn (in press) also suggest that it helps to reduce the stigma associated with certain policies - for instance, in the field of nuclear power - and highlights the reasons for public concern.

This kind of approach has the disadvantage that it is relatively labour-intensive, and that solutions are reached only after numerous revisions. However, of potentially greater concern is the appropriateness, or otherwise, of some of the assumptions underlying the procedure. A decision is usually redefined to ensure independence between different factors, allowing, in the simplest case, for the different factors to be added (Keeney, 1992). But there may be high correlations between factors, suggesting that at least some of them may be redundant: for example, the six factors designed to assess New Zealand government spending were found to capture the same underlying concern about the desirability of government policy (Kemp and Willetts, 1995).

Modelling preferences in this way may be appropriate for well-defined alternatives, but may be less satisfactory when we attempt to incorporate risk perceptions, where it is easy to imagine quite complex relationships between factors, so that efforts to capture these in quantitative models may lead either to quite complex solutions or else to gross simplifications. For instance, most applications require individuals to disregard regret and to treat ambiguity just like any other kind of uncertainty (Lichtenstein, Gregory, Slovic and Wagenaar, 1990; Gregory, MacGregor and Lichtenstein, 1992).

In practice, respondents may use a variety of non-compensatory decision rules which may not easily be incorporated into formal decision modelling. Process tracing may complement decision analysis by checking the extent to which information is used and in what order (Harte and Koele, 1995). This has shown that respondents vary the attention they give to the different factors and suggests a more evolutionary process for the decision problem, where structure and assessment occur simultaneously (Davey and Olson, 1994).

Decision analysts would normally structure the problem as they went along, but recent research has shown that weights are affected by the presentation of a decision hierarchy and the degree to which attributes are subdivided. This is worrying, although it has been argued that the differences are unlikely to dramatically change overall preferences (Borchercing and von Winterfeldt, 1988). However, others have suggested that decision analysis may bias responses by encouraging respondents to adopt a particular ethical position (Gregory, Keeney and von Winterfeldt, 1992; Fischhoff and Furby 1988). In addition, if the public's views of safety are not well-formed or well-articulated, decomposition of the decision into smaller units may run the risk of losing what little overall perspective respondents have (Fischhoff et al, 1978).

A more indirect method employs conjoint analysis. This approaches the problem by further simplifying the alternative offered to respondents. Rather than asking for weights or utilities directly, respondents are asked to rank, rate or choose between holistic alternatives. Weights and utilities are then inferred using regression analysis. This technique has been applied in market research (Lynch, Buzas and Berg, 1994), transport economics and environmental economics, and is increasingly being applied to health care (Ryan and Hughes, 1995; Vick...
and Scott, 1995, Propper 1988) However, if we were to apply it in its strictest sense, we would again face the problem of explicitly modelling the relationships between factors.

The alternative appears to be more along the lines suggested by Baron (1997): namely, to draw on some of the tools of decision analysis, but employ them in a rather looser way, primarily to help individuals structure the problem *en route* to arriving at a more considered, but essentially holistic, valuation. Thus instead of imposing rather fierce assumptions or restrictions and attempting to derive precise cardinal values and weights for each of the different dimensions, we might concentrate more on the ordinal ranking of policies and probe for the underlying values/beliefs. A range of in-built consistency checks might be used to help people refine their responses, and perhaps counteract the worst of the tendencies towards bias, inconsistency and insensitivity. But at present, as mentioned above, there are few examples of even small exploratory studies where such things have been attempted. Yet the difficulties and drawbacks of the various other methods which have been more widely tried and tested, and which have invariably fallen well short of the desired goal, suggest that this is the frontier which we should now be attempting to push back.
ACKNOWLEDGMENTS

This Review and Appendix was prepared under research grant No 3271/R73.04 from the Health and Safety Executive (with contributions also from: the Department of Environment, Transport and the Regions; Home Office; and HM Treasury) to the University of Newcastle, and sub-contracted to the Universities of York, Bangor and Sussex. Comments and other helpful input were provided by Mark Beatson, Jon Baron, Colleen Bowen, Baruch Fischhoff, George Lowenstein and Paul Slovic.
REFERENCES


APPENDIX A

Summaries of Selected Publications on Contingent Valuation (CV) and Other Direct Preference Elicitation Methods in the Fields of Health and Safety*

*In the interests of comparability, all monetary values in this Appendix have been updated and converted to £-Sterling, 1994. The updating and conversion was based on data given in the *International Financial Statistics Yearbook*, 1995. As an example of the procedure employed for updating and conversion, the figure of SEK 1,000,000 in 1986 prices reported in Persson and Cedervall (1991) was first updated to SEK in 1994 prices using the consumer price index and then converted to £-Sterling, 1994 on the basis of the 1994 average exchange rate, yielding a figure of £130,000. In addition, the summary table at the end of the Appendix also gives monetary values updated for increases in real income per capita under alternative assumptions concerning the income elasticity of the value of statistical life.

M.W. Jones-Lee (January 1998)
## APPENDIX A - CONTENTS

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Frankel, M. (1979) Hazard opportunity and the valuation of life. Mimeo, University of Illinois at Urbana-Champaign.</td>
<td>69</td>
</tr>
</tbody>
</table>


Summary of Contingent Valuation Estimates of the Value of Statistical Life

Acton’s survey, which was part of a larger study of the costs and benefits of different measures to reduce the risk of death by heart attack in the USA, involved a number of contingent valuation questions concerning willingness to pay for various programmes (such as a mobile cardiac ambulance unit) designed to reduce the probability of death should a heart attack occur. Three different samples were used. The first was a stratified random sample of 36 respondents, 18 men and 18 women, drawn from three communities in the Boston area during 1970. Acton describes these communities as “low income, working class”, “more affluent” and “affluent”. For this sample the questionnaires were administered by interviewers, although it is not clear whether these were professionally trained. The second sample of 21 trade union leaders was drawn from a group attending a 13-week course at Harvard Business School. This sample was given a brief explanation of the nature and purpose of the questionnaire prior to private completion. The third sample of 36 was drawn from a group of senior executives attending an advanced management course, also at Harvard Business School. This sample was self-selected from 160 executives who were asked to complete the questionnaire privately and without prior discussion.

While Acton’s study addressed a number of different issues, the most significant for present purposes concern

- willingness to pay for a heart attack programme that is essentially in the nature of a public good - i.e. one that is expected to save a particular number of lives in the local community.

- advice to a neighbour concerning the appropriate amount to pay for a programme which, should the neighbour suffer a heart attack, would reduce the risk that the attack would prove fatal

- and own willingness to pay for a similar programme.

In the public goods questions two different expected reductions in mortality were considered, namely the expected saving of 10 and 20 lives respectively in a community of 10,000 people. In the case of the “advice” and “own” willingness to pay questions, risk of heart attack and the reduction in risk of death conditional on heart attack were each varied across questions in such a way that the implied overall reductions in the probability of death during the coming year were as follows:

<table>
<thead>
<tr>
<th>Initial probability of death</th>
<th>Final probability of death</th>
<th>Reduction in probability of death</th>
</tr>
</thead>
<tbody>
<tr>
<td>$4 \times 10^{-3}$</td>
<td>$2 \times 10^{-3}$</td>
<td>$2 \times 10^{-3}$</td>
</tr>
<tr>
<td>$4 \times 10^{-3}$</td>
<td>$3 \times 10^{-3}$</td>
<td>$1 \times 10^{-3}$</td>
</tr>
<tr>
<td>$2 \times 10^{-2}$</td>
<td>$1 \times 10^{-2}$</td>
<td>$1 \times 10^{-2}$</td>
</tr>
<tr>
<td>$2 \times 10^{-2}$</td>
<td>$1.5 \times 10^{-2}$</td>
<td>$5 \times 10^{-3}$</td>
</tr>
</tbody>
</table>
As far as mean responses are concerned, with one or two exceptions these varied across questions in a more or less consistent manner (i.e. in the way predicted by theory). Thus, for example, in the “public goods” questions, mean willingness to pay to avoid 20 deaths exceeded that for the avoidance of 10 deaths in all three samples by a factor of about 3/2. Furthermore, in both the “advice” and “own safety” questions, willingness to pay for a reduction in risk from a given base level increased with the size of the reduction.

Intuitively, one might expect that larger reductions in risk from higher base levels would tend to elicit larger willingness to pay than smaller reductions from lower levels. While mean responses on the whole displayed this sort of pattern, individual responses were, in this respect more varied and were clearly regarded by Acton as being somewhat anomalous. However, more recent theoretical work suggests that these results may constitute less serious grounds for concern. In particular, results summarized in Jones-Lee (1989, p38). indicate that while willingness to pay for a $2 \times 10^{-3}$ reduction in annual probability of death from a base level of $4 \times 10^{-3}$ should exceed that for a $1 \times 10^{-3}$ reduction from the same base, nothing concrete can be said, a priori, about how such responses should relate to those for reductions from a different base (such as $2 \times 10^{-2}$). Acton appears to have hypothesized that willingness to pay would depend only upon the size of the risk reduction and, in particular, would be independent of the initial level of risk. Thus, in spite of the fact that two of his ‘own’ willingness to pay questions relate to a base risk of $4 \times 10^{-3}$ and two to a base of $2 \times 10^{-2}$, Acton clearly expected that the responses to all four questions would lie on a single increasing concave curve when plotted against the size of the risk reduction and, not surprisingly, he was disappointed in this expectation.

Regression analysis of the willingness to pay responses indicated that those factors that one might have expected to be the principal determinants of individual valuation of safety were, in fact, capable of explaining only a relatively small proportion of the overall variation in willingness to pay across individuals. In the main, coefficients of income and wealth variables were insignificant at the 5% level, while dummy variables for concern about heart disease, awareness of the heart attack ambulance programme, the respondent’s own state of health and headship of household tended to be significant.

Although values of statistical life implied by the responses to Acton’s survey vary across different questions and samples, they are on the whole very low in relation to estimates generated by the revealed-preference approach. In particular, responses to the “public goods” question (involving the saving of a specified number of lives in a community of 10,000) imply values ranging from $16,500 to $43,000, while responses to the “advice” questions imply values ranging from $12,100 to $47,000, and responses to the “own” safety questions imply values ranging from $7,400 to $43,000, all in 1970 prices. Converted to £-Sterling in 1994 prices, these figures become £41,100; £107,200; £30,200; £117,200; £18,400 and £107,200 respectively. Of all the estimates, Acton believed that most respondents “.....thought rather carefully about the [own safety] question that implies a value of $28,000 per expected life saved.....” – this was the question involving a $2 \times 10^{-3}$ reduction in the probability of death from a base of $4 \times 10^{-3}$. Updated to 1994 prices this becomes £69,800.

This study, described by its authors as “...only an experiment...,” intended to test the feasibility of the approach...” asked questions both about people’s perception of risk and their willingness to pay to reduce risk. The sample appears to have been self-selected and non-random and comprised 873 people attending Open Days at the UK Fire Research Station, Borehamwood, in 1972. Questionnaires were completed by the respondents themselves, presumably without the assistance of an interviewer.

The perception questions included such tasks as placing various activities in order of risk, comparing various risks with that of being killed or injured in a domestic fire, and estimating the proportion of fatal domestic fires in which there is more than one death. The results of these questions suggest that while people are, in the main, reasonably accurate in their ranking of risks, their perceptions of the numerical magnitude of these risks are not so accurate when compared with historical frequencies.

The questionnaire contained three different questions concerning valuation of safety. However, one of these involved a trade-off between time and safety and the responses will not be reported here, save to note that as part of the question respondents were required to estimate the incremental risk of death when overtaking on a road with two-way traffic. Responses tended to confirm the earlier observation that most people are poor at estimating risks in numerical terms: the mean perceived risk was $1.4 \times 10^{-4}$ and the median $8 \times 10^{-5}$ whereas the “actual” risk, based on historical frequencies, is $1.4 \times 10^{-8}$.

The two contingent valuation questions concerned (a) willingness to pay for a domestic smoke detector which, it was asserted, would reduce by 90% the (then) existing annual risk of death in a domestic fire of 1 in 80,000 per person, and (b) willingness to pay for hypothetical “safe” cigarettes. In the case of the smoke detector question it was assumed that responses reflected willingness to pay for own safety as well as that of other members of the household and, accordingly, an average willingness to pay “per person at risk” was computed. This, together with the risk reduction given in the question, implied a value of statistical life of £55,000 in 1972 prices. Interestingly, however, responses were highly right-skewed with a value of statistical life of only £20,000 implied by the median responses. Updated to 1994 prices, these figures become £365,500 and £132,900 respectively.

Finally, the value of statistical life implied by the mean response to the “safe” cigarettes question was £17,200, again in 1972 prices. Updated to 1994 prices, this becomes £114,300.

Unfortunately, no questions concerning age, income etc. were asked, so that it was not possible to attempt to explain variations in willingness to pay by means of regression analysis.

This study, like that by Melinek et al. (1973), was not intended as a definitive exercise in the estimation of values of safety, but rather as an example of the way in which one might proceed to obtain empirical estimates and as a means of exploring the feasibility of the contingent valuation method and the possible order of magnitude of the results that it might be expected to produce. The study also included an extensive discussion of the scientific status of questionnaire-based exercises in this context.

Basically, the questionnaire contained just two types of contingent valuation question, each involving various increases and decreases in risk from a given base level. The first type of question concerned willingness to pay or required compensation (in the form of maximum acceptable increases or minimum acceptable reductions in an air fare) for variations in air travel safety from a base of $4 \times 10^{-3}$ per flight. The second type of question asked about the premia or discounts on a standard house purchase price that would just induce the respondent to live in areas with varying levels of environmental pollution when the sole effect of the latter was to increase or decrease life expectancy by specified amounts.

The sample was small, self-selected and non-random, comprising 31 academics, researchers and public sector employees, and the survey, conducted by postal questionnaire, was carried out early in 1975. The majority of respondents provided answers that seemed, *prima facie*, to meet minimal conditions of consistency. For the air safety questions the implied value of statistical life in 1975 prices was £3,100,000 for a “one-off” air journey involving self-only (as opposed to self plus family) and £2,500,000 for a journey that would have to be undertaken weekly for a year, again on a self-only basis. Updated to 1994 prices, these figures become £11,200,000 and £9,000,000. Distributions of responses were bi-modal and right-skewed so that median responses were some 10% - 25% below means.

Responses to the questions concerned with environmental pollution implied marginal rates of substitution of wealth for life expectancy of £900 per year for self-only and £1,790 per year for self plus family. While it is tempting to suppose that values of statistical life could be inferred directly from these estimates of willingness to pay for increased life expectancy, there are two reasons why this is not so. In the first place, the questions concerning environmental pollution were worded in such a way as to imply that the variation in life expectancy would reflect *simultaneous* variations in the probability of death for each of a number of future years, whereas inference of a value of statistical life would require that we consider a variation in life expectancy due to a *ceteris paribus* change in the probability of death for the coming year only. The only circumstance in which this difficulty would disappear seems to be that in which individuals value a unit increase in life expectancy *independently* of the pattern of age-specific probabilities of death that produce that increase. One could then infer willingness to pay for a *ceteris paribus* reduction in the probability of death for the coming year from the willingness to pay for increased life expectancy, since it is straightforward to show that, with $p_1$ small, \( \frac{d\mu}{dp_1} = \mu \) where \( \mu \) denotes life expectancy and \( p_1 \) denotes the probability of death during the forthcoming year. Even if this condition were fulfilled, however, there would remain a further difficulty, namely that the smallest increase in life expectancy referred to in the study was one year. Now, in order to produce such an increase, the *ceteris paribus* reduction in \( p_1 \), given that \( \frac{d\mu}{dp_1} = \mu \) would have to be approximately equal to \( 1/\mu \) which, being many multiples of most people’s current level of \( p_1 \), would simply not be credible.

It is surprising that this study has not been more widely discussed in the literature, since it represents an extremely thorough attempt to assess the workability of the contingent valuation approach as a means of arriving at reliable and valid estimates of individual valuation of safety.

The questionnaire contained a number of different questions which can be categorized, broadly speaking, as follows:

- Questions designed to elicit respondents’ numerical estimates of various risks, including risk of death in a domestic fire, risk of death in a hotel fire, risk of death in a road accident and risk of death by electrocution.

- Questions designed to elicit respondents’ estimates of the percentage reduction in each of the above risks afforded by various safety devices such as domestic smoke detectors and safety tyres.

- Questions designed to establish respondents’ willingness to pay for the various devices referred to above. In the case of fire safety, questions involving both private and public provision were included.

- In-depth questions designed to check on the quality of respondents’ understanding of the probabilistic concept of risk, to check as far as possible on the veracity of valuation responses, and to explore the general nature of attitudes to safety. These questions were put to a subsample of those who had answered the earlier categories of question.

The quota sample of 325 respondents, designed to be representative of the overall population in terms of age, gender and socioeconomic group, was drawn from the London area. Interviews were carried out in the respondents’ own homes by members of Research Surveys of Great Britain Ltd during 1978 and 1979. This professional survey organization had also refined the questionnaire following extensive piloting by the author. Interviews lasted between 30 minutes and one hour.

The questionnaire design included many novel features, but one in particular deserves special mention, namely the procedure for conveying information about probabilistic risks. Briefly, three risks (being struck by lightning, being killed as a motorcyclist and being killed as a pedestrian) were described in terms of the number of deaths per year in a group the size of the population of a large town such as Birmingham (which happened, at the time of the study, to have about a million inhabitants). This information was supplemented by a pictorial display using a “thermometer chart”. In all subsequent questions concerning risks, respondents were asked to express their answers in similar terms and, if they so chose, to do so using the pictorial display. The in-depth questions suggested that respondents were comfortable with and understood the concept of risk explained in this way. In addition, Maclean notes that:

‘The ease with which most people were able to express their estimates of the levels of risk helped to dispel any preconceptions of difficulties in dealing with small probabilities’ (Maclean, 1979 p10)
The results of the survey are many and varied but the following seem to be the most significant for present purposes.

- Respondents appeared in the main to be able to comprehend the concepts involved in the questions and were willing to provide apparently consistent answers.

- Respondents' perception of the rank ordering of various risks accorded well with the "objective" ordering based on relative frequencies. However, subjective perceptions of the numerical magnitude of these risks varied widely.

- Roughly 50% of respondents were provided with "objective" risk data after they had supplied their own subjective estimates but before answering the contingent valuation questions. However, no significant difference was found between the valuation responses of these respondents and the remainder of the sample. This finding was confirmed by responses to the in-depth questions which indicated that subjective "feelings" for a level of risk were not influenced by information concerning the "objective" level.

- Responses to contingent valuation questions produced a wide variety of estimates of the value of statistical life. However, while there are doubts about the procedure employed to arrive at some of these estimates, it appears that the value of statistical life based on willingness to pay for domestic fire safety was correctly computed and this was estimated to be £1,400,000 in 1979 prices. Updated to 1994, this figure becomes £3,560,000.

- While it would appear that there was no significant relationship between valuation responses and gender or social class, some questions did produce a significantly decreasing relationship with age.

This study, like Maclean’s, has many interesting features and produces a number of useful insights. The questions concerned willingness to pay for air safety and willingness to pay, or requirement of compensation, for variations in life expectancy.

The sample of 169 was made up of academics (72%) and middle-rank executives attending a Master of Business Administration programme (28%). As such, the scenarios with which respondents were confronted, as well as the nature of the sample, were very similar to those in Jones-Lee (1976).

The first of two air safety questions asked about willingness to pay to eliminate a risk of $1.5 \times 10^{-6}$ of a fatal accident on a single flight. The mean response was $18.33 and the median $4.45$, implying values of statistical life of $12,000,000$ and $3,000,000$ respectively, both in 1979 prices. Converted to £-Sterling in 1994 prices, these figures become £19,000,000 and £4,800,000 respectively. For this question, 23% of respondents indicated that they would be willing to pay nothing for the safety improvement and 12% reported a willingness to pay of £100 or more. Asked to rank the risk of $1.5 \times 10^{-6}$ on a scale from 1 (“negligible”) to 6 (“very high”), 74% of respondents reported that they regarded the risk as “negligible”.

A second air safety question asked about willingness to pay to eliminate a risk of $1 \times 10^{-3}$ on a single flight. In this case the median response was £50.37 (no mean is reported), implying a value of statistical life of £50,000. Converted to £-Sterling in 1994 prices, this figure becomes £79,800. For this question, over 50% of respondents regarded the risk as “high” or “very high”.

The questions dealing with variations in life expectancy asked about willingness to pay for one-year and five-year increases in life expectancy, willingness to pay for a certain increase in length of life by one year, and finally requirement of compensation for the loss of one year and five years of life expectancy. The responses to all of these questions varied very widely indeed, with over 30% of the sample indicating unwillingness to pay anything for the various increases in life expectancy and a substantial proportion (15% - 30%) reporting a willingness to pay of over £5,000.

The mean willingness to pay for a one-year increase in life expectancy was just over £1,000 (which is broadly comparable with the result for a similar question reported in Jones-Lee, 1976), although it should be noted that the median response was a mere £5.33. Not surprisingly, willingness to pay for the certain extra year of life was substantially higher, with a mean of over £1,800 and a median of £500. Compensation required for the one-year loss of life expectancy was also very substantial, only 3% of the sample indicating that they would require nothing and 26% stating that over £100,000 would be needed. In this case the mean response was £45,000 and the median £30,000. In general, the questions involving five-year variations in life expectancy elicited responses that moved in the “right” direction in relation to their one-year counterparts. As was noted in the summary of Jones-Lee (1976), it seems inappropriate to attempt to derive values of statistical life from the responses to the questions involving variations in life expectancy.

Undoubtedly, the two most significant features of the results of Frankel’s study are the enormous disparity between the values of statistical life implied by the responses to the two air safety questions and the large proportion of the sample reporting unwillingness to pay anything for an increase in life expectancy (44% for the one-year increase and 33% in the five-year case). As far as the disparity between the values of statistical life are
concerned, this could be taken as evidence of the inherent unreliability of the contingent valuation approach in general and one has to admit that these results do provide cause for some anxiety on this score.

It should be borne in mind, however, that the risk reductions involved in the air safety questions were, respectively, very small and very large indeed. In the case of the very small reduction, any estimate of the value of statistical life based on stated willingness to pay for such a reduction will be extremely sensitive to even small variations in the amount concerned (in particular, with a risk reduction of $1 \times 10^{-6}$, a $1$ variation in the mean response will produce a $666,667$ change in the value of statistical life). If one adds to this the fact that the typical response to this question involved a very small amount of money relative to the respondent’s income - so that reported willingness to pay may not have been particularly carefully thought out - then it would seem reasonable to conclude that one should not expect very reliable estimates of the value of statistical life from questions involving such small changes in risk. For rather different reasons, estimates derived from questions involving very large risk reductions are also likely to be of doubtful reliability. For example, a risk reduction of $1 \times 10^{-3}$ represents a very substantial proportion of the annual risk of death faced by most people and one suspects that many respondents would find it difficult to imagine that such a large reduction could actually be “delivered” unless there were, in addition to any sum paid for the reduction, very substantial costs of discomfort or inconvenience - as in the case of giving up smoking or undergoing radiation treatment or chemotherapy for cancer. In other words, a situation in which one might actually be able to purchase, say, a 25% or even 50% improvement in one’s survival prospects for the coming year without incurring substantial discomfort or inconvenience costs is so far outside most people’s experience that, again, responses to questions involving such large changes should probably not be taken too seriously.

The long and short of all this is that Frankel’s results, rather than pointing out a fundamental flaw in the contingent valuation approach, are probably better viewed as highlighting the importance of posing questions that are believable, relate to familiar situations, and involve choices or decisions that respondents will “take seriously”.

The zero responses to the increased life expectancy questions are, *prima facie*, more puzzling and disturbing. Frankel offers various possible explanations for these results. For example, he suggests that a proportion of people may believe that “……..their life spans, though unknown, are essentially predetermined by spiritual or mechanistic forces……”.

Another effect that Frankel conjectures may have been at work is “…..the belief that one’s longevity is a very chancy thing, and that an expected year more or less doesn’t count for much”. Finally, Frankel notes that, in written comments, several respondents referred to their concern about the possibility of poor health during any period of extended longevity.

This study was designed to shed light on a range of issues concerning individual perception and valuation of transport safety. The study involved a number of contingent valuation questions concerning willingness to pay, or requirement to compensation, for various kinds of safety effect (including changes in own safety, other people’s safety, effects that were in the nature of private goods, those that were publicly provided and so on). However, the results that are most directly comparable with those of the other empirical studies reviewed in this Appendix were the responses to questions concerning (a) willingness to pay for a car safety feature that would provide various specified reductions in the annual probability of death in a car accident and (b) willingness to pay, or requirement of compensation, for variations in coach safety on a foreign trip. In these questions variations in risk were expressed in multiples of 1 in 100,000, respondents having been told that the annual risk of death in a car accident in the UK for a car driver was at the time of the study, on average, about 10 in 100,000. In addition, all statements about risk were accompanied by a visual display in which the appropriate number of squares had been “blacked out” on a sheet of graph paper containing 100,000 squares.

The study was commissioned by the UK Department of Transport, following a one-year feasibility exercise, and the survey was carried out during June and July 1982 by National Opinion Polls Ltd (NOP) using a questionnaire designed and piloted by the authors and further piloted and refined by NOP. The three-stage stratified random sample was drawn from 93 parliamentary constituencies in England, Scotland and Wales and produced 1103 full and 47 partial interviews. The response rate was 67% and interviews, conducted by professional interviewers, on average took about 45 minutes to complete. A follow-up survey of 210 of the respondents to the main study was conducted to test for the temporal reliability of responses.

The main findings of the study were as follows.

- A substantial majority of respondents were willing and able to provide answers to the various questions put to them.

- While a proportion of respondents gave apparently inconsistent answers to some of the questions and experienced difficulty with the concepts involved, it was the authors’ opinion that the balance of the evidence from the survey supported the view that the quality of perception of transport risks, ability to process probability information and the veracity of responses were, on the whole, sufficient to justify the inference of at least broad orders of magnitude of the value of statistical life for transport risks.

- The values of statistical life implied by mean responses to the car and coach safety questions referred to above (with a few suspect outlying responses removed) ranged from £1,200,000 to £2,000,000 in 1982 prices, while values based on median responses ranged from £500,000 to £1,200,000. Given these results, the authors argue for a value of statistical life for “self-only” transport risks of at least £500,000, with a value closer to £750,000 warranted by consideration of median responses and about £1,500,000 by (trimmed) means. Updated to 1994 prices, these figures become £887,000; £1,330,000 and £2,660,000 respectively.

- Regression analysis indicates that individual valuation of safety is significantly affected by income and age, with the coefficient of a linear income variable.
implying an income elasticity of the marginal rate of substitution of wealth for risk of death of about 0.3, and the coefficients of linear and quadratic age variables implying that the marginal rate of substitution follows a clear “inverted-U” life-cycle.

- Under the willingness to pay definition, the value of avoidance of a serious injury (where seriousness was construed in terms specified by the respondent) was at least one-hundredth of the value of statistical life.

While the results of this study do not permit the inference of unambiguous values of
statistical life, they do bear directly upon the properties of the individual safety valuation
function or, more accurately, upon the *multi-period counterpart* to this function. In
addition, Smith and Desvousges' results have some features that are strikingly similar to
the (unpublished) results of various pilot-scale studies conducted by the UK Health and
Safety Executive. It therefore seems appropriate to include a brief summary of the main
findings of the study of Smith and Desvousges in this Appendix.

Smith and Desvousges' principal purpose was to test various "standard" hypotheses
concerning the relationship between individual willingness to pay and the size and nature
of the risk reduction concerned. For example, simpler theoretical models predict that
willingness to pay will be an increasing, strictly concave function of reduction in the
probability of death during a forthcoming period and also that willingness to pay for a
given reduction in this probability will be an increasing function of its initial level (see,
for example, Jones-Lee, 1976; or Weinstein *et al*, 1980). Another implication of
standard theory is that the amount that an individual would be willing to pay to reduce the
probability of death during a forthcoming period from $p_a$ to $p_b$ (with $p_b < p_a$ ) will be
identical to the sum he would pay to prevent the probability rising from $p_b$ to $p_a$, other
things equal. Finally, since their main interest was in risks from hazardous wastes, Smith
and Desvousges wished to examine the extent to which standard hypotheses - which
typically apply to variations in the *annual* risk of death - stand up in situations in which
risks vary over a very much more protracted period. In fact, Smith and Desvousges
considered variations in the probability that hazardous wastes would cause premature
death within a 30-year period. (It is essentially for this reason that one cannot derive
unambiguous estimates of the value of statistical life from their results).

The survey and questionnaire design were relatively sophisticated and involved extensive
pre-testing, in-depth discussion with members of the public, practice sessions for
interviewers and so on. The stratified cluster sample of about 700 was drawn from the
Boston area and produced over 600 fully completed interviews. The sample was randomly
partitioned into eight subsets (referred to as "design points"), while combinations of base
risk and risk reduction (or increase) were varied across design points in order to test the
hypotheses described above.

All statements about risk were presented in the form of a risk $R$ of being exposed to
hazardous waste during the 30-year time horizon and a separate conditional risk $q$ of
death from the exposure. This ensured that the separate risks $R$ and $q$ were, in most cases,
large enough to be representable in a "pie-chart" visual display that accompanied the
questions.

In essence, the main findings of the survey were as follows.

- As in other questionnaire studies, the vast majority of respondents were willing
  and able to answer the questions asked of them.

- Within each design point, mean willingness to pay was - as predicted by theory -
  an increasing function of the size of risk reduction from a *given* base risk.
  However, while in some cases the function was concave, in others the evidence
  was consistent with linearity or even convexity.

- Far from increasing with the level of base risk, mean willingness to pay for a
  given risk reduction showed a marked tendency to *fall*. In addition, and somewhat
more disturbingly, design points with larger risk reductions did not appear to elicit correspondingly higher mean or median willingness to pay.

- The mean willingness to pay to reduce risks from $p_a$ to $p_b$ (with $p_b < p_a$) was uniformly larger than mean willingness to pay to avoid an increase from $p_b$ to $p_a$.

- When the non-zero willingness to pay responses, divided by the corresponding variation in risk, were regressed on various explanatory variables with a semilogarithmic specification, consistently significant income coefficients were obtained. These coefficients imply income elasticities of between 0.18 and 0.24 which, are very similar to the estimates reported by Blomquist (1979) and Jones-Lee et al (1985).

The findings of this study are therefore, to say the least, mixed, both in relation to conventional wisdom and to the results of other empirical studies. The authors offer various possible explanations for the major discrepancies, including the fact that their study dealt with risks that are relatively poorly understood, involve the possibility of painful, lingering death and were explicitly described in the study as applying over a protracted period. In addition, they suggest that various of the phenomena identified by work in experimental psychology - such as "anchoring", and subjective overestimation of low "objective" probabilities - may go some way towards explaining the more puzzling relationships between responses from different design points.

In fact, Smith and Desvouis's results taken as a whole probably represent somewhat less serious cause for concern than might at first appear to be the case. In the first place, while simpler and more restrictive theoretical analyses certainly do predict strict concavity of the willingness to pay versus risk reduction relationship, there are grounds for expecting that this relationship may well be approximately linear or even, in some cases, locally convex, as noted in Jones-Lee (1989) Ch3. Furthermore, Dehez and Drèze (1982) have shown that quite plausible assumptions concerning the properties of life insurance and annuity contracts - typically ignored in the simpler models - can produce predictions about the relationship between willingness to pay and base levels of risk that are entirely consistent with Smith and Desvouis's findings.

Nonetheless, it must be admitted that two aspects of the results of this study defy explanation in terms of the conventional conceptual apparatus of economic analysis. These concern the fact that, as one moves from one design point to another, larger reductions in risk do not appear to elicit larger willingness to pay and that mean willingness to pay for a given risk reduction significantly exceeds mean willingness to pay to avoid the corresponding increase in risk. In order to make sense of these results one probably has to appeal to the ideas of cognitive psychology (such as anchoring and framing) rather than economics.
The purpose of this study was to test various hypotheses concerning individual willingness to pay, and willingness to accept compensation, for variations in health risks. The hypotheses were drawn not only from the economics literature, but also from cognitive psychology. Denoting the variations in two independent health risks by $\alpha$ and $\beta$ and the compensating variation in wealth (i.e. willingness to pay or willingness to accept compensation) by $v$, the hypotheses generated by the economics literature are then:

- $\frac{\partial v}{\partial \alpha} > 0, \frac{\partial v}{\partial \beta} > 0, \frac{\partial^2 v}{\partial \alpha^2} < 0, \frac{\partial^2 v}{\partial \beta^2} < 0$

- $\alpha = \beta = \gamma \Rightarrow \frac{dv}{dy} = \frac{\partial v}{\partial \alpha} + \frac{\partial v}{\partial \beta} > 0$

- If $\alpha$ and $\beta$ are non-marginal and negative (i.e. risk reductions) then:

$$v(\alpha, \beta) > v(\alpha, 0) + v(0, \beta)$$

Where $v(\alpha, \beta)$ denotes the compensating variation for simultaneous reductions in the two risks. Though this is not spelled out in the paper, it would seem that if both risk reductions are normal goods then:

$$v(\alpha, \beta) < v(\alpha, 0) + v(0, \beta)$$

In the case of hypotheses suggested by the cognitive psychology literature, these are as follows:

- For any given reduction in risk, individuals place an added value on the complete elimination of risk (a) because this reduces anxiety and any costs associated with thinking about a probabilistic outcome and (b) because individuals over-estimate and/or overweight initial small probabilities but accurately assess the terminal risk of zero.

- Individuals weigh increases in risk from a given reference point very much more heavily than they weigh reductions in risk of the same magnitude from the same reference point.

In order to test these hypotheses, the authors conducted a contingent valuation survey involving willingness to pay (WTP) and willingness to accept (WTA) premia (or discounts) on the standard purchase prices of a fictitious but professionally prepared and labelled insecticide spray and lavatory bowl cleaner, in order to effect various reductions (or accept increases) in each of two risks associated with each of the two products. For respondents from households with children under the age of five, the risks associated with the insecticide were inhalation and child poisoning, while those associated with lavatory bowl cleaner were eye burns and child poisoning. For respondents from households
without children under the age of five, the insecticide risks were inhalation and skin poisoning, while those associated with the lavatory bowl cleaner were chloramine gassing and eye burns. Each respondent considered only one of the two fictitious products and was told its current price ($10 per bottle for the insecticide and $2 for the lavatory bowl cleaner) as well as the current risk level of 15 injuries of each of the two types per 10,000 bottles.

The survey, carried out sometime prior to 1987, was conducted in a shopping mall and hardware store in Greensboro, North Carolina by professional interviewers from a market research firm, supervised by the authors. The sample was drawn from a representative mix of consumers and, in the case of the insecticide spray, comprised 672 respondents without children under the age of five and 113 with children under the age of five, while those who answered the questions concerning the lavatory bowl cleaner consisted of 551 respondents without young children and 183 with.

For each of the two fictitious products, respondents answered six CV questions. In particular, denoting by $(\alpha, \beta)$ a change in the first risk by $\alpha$ in 10,000 bottles and a change in the second risk by $\beta$ in 10,000 bottles, the structure of the six questions and the corresponding hypotheses tested were as follows:

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Changes in Risk $(\alpha, \beta)$</th>
<th>Hypotheses Tested</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(-15, 0)</td>
<td>Positive valuation of risk reduction; certainty premium</td>
</tr>
<tr>
<td>2</td>
<td>(0, -15)</td>
<td>Positive valuation of risk reduction; certainty premium</td>
</tr>
<tr>
<td>3</td>
<td>(-15, -15)</td>
<td>Positive, but diminishing valuation of risk reduction, certainty premium</td>
</tr>
<tr>
<td>4</td>
<td>(-5, -5)</td>
<td>Positive valuation of risk reduction</td>
</tr>
<tr>
<td>5</td>
<td>(-10, -10)</td>
<td>Positive but diminishing value of risk reduction</td>
</tr>
<tr>
<td>6</td>
<td>(+1, +1)</td>
<td>Negative valuation of risk increases, WTP/WTA disparity</td>
</tr>
</tbody>
</table>

More specifically, questions 1 to 5 provide direct tests of positive WTP for risk reductions, while question 6 provides a direct test of positive WTA for an increase in risk. In addition, question 6 gives a direct test of the WTP/WTA disparity in relation to a given reference risk. Questions 3,4, and 5 together provide a test of diminishing incremental valuations of additional reductions in risk, while questions 3 and 5 provide a test of whether there is a certainty premium associated with reducing both risks to zero that exceeds the effect of diminishing incremental valuations. Finally, questions 1,2 and 3 provide a test of the additivity (or otherwise) of $v(\alpha,0)$ and $v(0,\beta)$ in relation to $v(\alpha,\beta)$.

The main findings of the study were as follows:

- The responses to questions 1 to 5 gave a positive and statistically significant valuation of risk reduction.
At both the aggregate and individual level, the responses to questions 3, 4 and 5 lend support to the hypothesis of diminishing incremental willingness to pay for additional risk reduction. For example, the mean willingness to pay for the (-5, -5) risk reduction was substantially and significantly larger than the difference between the mean willingness to pay for (-10, -10) and (-5, -5).

- In the case of the relationship between \( v(-15, -15) \) on the one hand and \( v(-15, 0) + v(0, -15) \) on the other, the hypothesis of equality could not be rejected at the 5% level.

- There was strong evidence in support of a certainty premium in that in most cases the difference between the mean willingness to pay for (-15, -15) and (-10, -10) substantially exceeded that between (-10, -10) and (-5, -5) and indeed mean willingness to pay for (-5, -5) itself. This finding is supported at the level of individual responses.

- The hypothesis of a significant WTP/WTA disparity was dramatically supported by the evidence, in that between 61.5% and 77.2% of respondents indicated that they would not be willing to accept the (1, 1) increase in risk at any discount on the purchase price of the product concerned.

Apart from a question concerning the wage premia required to compensate for increased job risk posed in the survey by Prescott-Clarke (1982) - which was insufficiently specific about the nature of the risk to permit inferences concerning marginal rates of substitution of wealth for risk, - to the best of our knowledge, this is the first thoroughgoing contingent valuation study of job risks, as such.

In addition to the usual factual and demographic questions, the study sought to determine (i) the increase in annual wages that would be required to induce the respondent to work at his or her existing job if the risk of accidental death at work were one step higher than the current risk on a ten-step “ladder” of risks and (ii) how much of his or her current annual wages the respondent would be prepared to forego if their current job risk were to be moved one step lower on the ladder. Willingness to Pay (WTP) and Willingness to Accept (WTA) amounts were elicited using a payment card with 37 boxes ranging from $0 to more than $6000. While respondents were not informed about the actual increment in risk as one moved up each step on the risk ladder (the figure was, in fact, about 1 in 4000 per annum), seven representative occupations ranging from School Teachers to Lumberjacks were located at the appropriate points on the ladder and respondents were then invited to pick the point on the ladder that most closely corresponded to their perception of their own job risk.

In the case of the WTA question concerning a move up the ladder, respondents who initially located themselves at the top of the ladder were asked to answer as if they were in fact initially one step below the top, while in the case of the WTP question concerning a move down the ladder, respondents who initially located themselves at the bottom of the ladder were asked to answer as if they were in fact initially one step above the bottom. In this restricted sense the study was based upon subjective perceptions of “own” job risk, though to the extent that estimation of marginal rates of substitution was based upon the “objective” risk increment of 1 in 4000 per annum for each step up the ladder, objective risks did enter into the calculation. In addition to providing information concerning their perceptions of own job risk by means of the risk ladder, respondents were also presented with a list of 13 major causes of death at work (such as motor vehicle accident, electrocution or explosion) and asked to rank the likelihood of each occurring to them on a five point Likert scale running from 1 (“Could Never Happen”) to 5 (“Most Likely to Happen”). An overall job risk index (RISK 1) for each respondent was then computed as the average of these scores over the 13 causes of death.

The study, involving a nationally representative postal survey, was conducted during the summer of 1984. Survey materials were posted to (i) a simple random sample of 3000 U.S. households and (ii) 3000 additional households randomly sampled from 105 counties that had disproportionately large concentrations of high-risk industries. Of the 6000 questionnaires mailed, 749 (12.5%) were returned as undelivered and 2103 were returned in completed form, giving a net response rate of about 40%. Of the completed questionnaires 872 from retired or unemployed individuals were excluded, as were those from 32 individuals who did not complete the contingent valuation questions. An additional 338 individuals who did not supply information concerning labour earnings or initial level of job risk were also excluded, so that the final sample size was 861.

As is often the case with contingent valuation questions, responses to both WTP and WTA questions were highly right - skewed with means exceeding medians and a substantial number of zero bids (47.4% for the WTP question and 23.2% for the WTA). The mean response to the WTP question (presented to 50% of the sample) was $665 for a one-step move down the risk ladder, while that for the WTA question (presented to the
balancing 50% of the sample) was $1705 for a one-step move up the ladder. Given that a one-step move on the risk ladder entailed a change in the annual risk of a fatal accident at work of 1 in 4000, these mean WTP and WTA responses give values of statistical life for job-related risks of $2,660,000 and $6,820,000 respectively in 1984 prices. Converted to £-Sterling in 1994 prices, these figures become £2,480,000 and £6,350,000 respectively.

Finally, regression analysis indicated that both the WTP and WTA contingent valuation responses were significantly positively related to annual earnings and the perceived initial level of job risk as measured by the overall job risk index, RISK I

The purpose of this study was to obtain estimates of WTP-based values for the prevention of insecticide spray poisoning, taking explicit account of people's altruistic willingness to pay to reduce risks to other people (a) within their own State and (b) in the whole of the rest of the U.S.A. The contingent valuation questions presented to respondents who did not have children under five years of age concerned (a) additional willingness to pay, over and above a specified market price, for an insecticide spray with reduced risks of inhalation poisoning and skin poisoning for members of the respondent's household and (b) willingness to contribute to an advertising campaign that would reduce the risks of insecticide spray inhalation and skin poisoning for the population of North Carolina. Respondents were then asked about their additional willingness to pay - over and above the latter amount - to extend the advertising campaign to the entire population of the U.S.A. In the case of respondents with children under five, the questions concerning willingness to pay to reduce the risk of skin poisoning (both for own household members and for the populations of North Carolina and the U.S.A.) were replaced with questions concerning willingness to pay to reduce the risk of child poisoning as a result of drinking insecticide.

The survey, carried out sometime prior to 1988, was conducted in Greensboro, North Carolina by professional interviewers from a North Carolina marketing firm, monitored by two graduate students. The locale from which respondents were recruited is widely used in national marketing studies for major consumer brands in view of its representative population mix. The sample included 672 respondents without children under five and 113 respondents with children under five. To be eligible for inclusion in the sample, respondents had to be either over age 21 or else over age 20 and not a student.

Prior to answering the CV questions, respondents were asked to examine a professionally prepared and labelled (but fictitious) insecticide spray and were then asked a series of questions about precautionary behaviour. The CV questions concerning willingness to pay to reduce risks to members of the respondent's own household then followed. After questions about willingness to pay for complete elimination of insecticide poisoning risks, respondents without children under five were asked about their additional willingness to pay per bottle of insecticide spray - over and above the "baseline" price of $10 - to reduce each of the risks of inhalation and skin poisoning from 15 to 10 per 10,000 bottles used. Throughout the analysis reported in the paper, this additional willingness to pay is treated as being that for a $5 \times 10^4$ reduction in the risk of an injury "pair" comprising one case of inhalation and one case of skin poisoning (rather than as a $10 \times 10^4$ reduction in the risk of suffering one or other type of poisoning.

Provided that the injuries were regarded as being (roughly) equally undesirable then the additional willingness to pay per injury avoided will be approximately half the additional willingness to pay for the injury pair. In fact, the mean additional willingness to pay for a $5 \times 10^4$ reduction in the risk of the injury pair inhalation/skin poisoning was $1.04 while that for the inhalation/child poisoning was $1.84. Given the risk reduction of $5 \times 10^4$, these figures translate into values for prevention of the respective injury pairs of $2080$ and $3680$, presumably in approximately 1987 prices. Converted to £-Sterling in 1994 prices, these figures become £1,770 and £3,130.

Having answered the "private" risk reduction CV questions, respondents were then asked about their willingness to contribute to a non-profit-making advertising campaign which would reduce the number of inhalation poisoning and skin poisoning (or inhalation and child poisoning) pairs from 15 per 10,000 bottles to 10 per 10,000 bottles for the coming year for the whole of North Carolina. Respondents were told that in view of the
fact that there are 2 million households in North Carolina, this would reduce the number of injury pairs from 3,000 to 2,000 for the State as a whole. Respondents were then asked about their additional willingness to pay to extend the advertising campaign from North Carolina to the whole of the U.S.A. In the case of the inhalation/skin poisoning pair, 57% of respondents indicated a willingness to contribute to the North Carolina advertising campaign, while 14% indicated a willingness to make an additional contribution to extend the campaign to the whole of the U.S.A. For the sample as a whole (including those unwilling to contribute) mean willingness to contribute to the North Carolina advertising campaign was $5.01 while the additional willingness to pay to extend the campaign to the whole of the U.S.A. was $1.72. In the case of the inhalation/child poisoning pair the corresponding figures were 79%, 21%, $9.06 and $2.39 respectively. Multiplying mean willingness to contribute by the number of households in North Carolina (or where appropriate, in the rest of the U.S.A.) and dividing by the number of injury pairs prevented gives collective altruistic willingness to pay per inhalation/skin poisoning injury pair prevented of $10,000 for North Carolina and $3070 for the rest of the U.S.A., while for the inhalation/child poisoning pair the corresponding figures are $18,100 and $4260 respectively. Converted to £-Sterling in 1994 prices, these figures become £8,500; £2,160; £5,400 and £3,630 respectively.

Finally, regression analysis of the altruistic willingness to pay responses indicated that private willingness to pay for household risk reduction and household income had the most consistently significant (positive) impact on altruistic willingness to pay for others' safety and that other demographic variables had little, if any, impact.

The purpose of this study was to test various hypotheses that have been proposed in the literature concerning attitudes towards fairness in the distribution of risks and benefits. In order to introduce these hypotheses it will be convenient to consider a two-person world and to denote by \((p_1, p_2)\) a situation in which the two individuals face statistically independent risks of death of \(p_1\) and \(p_2\) respectively. The main hypotheses considered by Keller and Sarin are then:

- **Ex ante equity - proneness** which entails that \((p_1, p_2)\) will be preferred to \((p_1 + \varepsilon, p_2 - \varepsilon)\) whenever the magnitude of the difference between \(p_1\) and \(p_2\) is smaller than the magnitude of the difference between \(p_1 + \varepsilon\) and \(p_2 - \varepsilon\) (see, for example, Keeney, 1980b).

- **Ex post equity - proneness** which entails that a 50/50 chance of \((0,0)\) and \((1,1)\) will be preferred both to a 50/50 chance of \((0,1)\) and \((1,0)\) and to \((0.5, 0.5)\) (see, for example, Fishburn, 1984).

- **The distribution of benefits can, to some extent, offset inequalities in the distribution of risks** which entails that with \(p_1 > p_2\) and \(\varepsilon > 0\), \((p_1 + \varepsilon, p_2 - \varepsilon)\) may be preferred to \((p_1, p_2)\) if the move from \((p_1, p_2)\) to \((p_1 + \varepsilon, p_2 - \varepsilon)\) benefits person 1 in some other way.

In order to test these and other hypotheses, Keller and Sarin presented 106 graduate students at the University of California, Irvine, with a number of scenarios involving pairwise choices. The subjects were divided into two equal-sized groups, the first of which was simply asked to express a preference in each of the pairwise choices while the second was asked to compare the options on the basis of fairness (which was, apparently, given no formal definition as far as subjects were concerned). The various scenarios were as follows:

**Scenario 1**, which was intended to test for **ex post equity - proneness**, involved the choice between (a) selecting one from a group of 100 islanders (presumably at random) who would then be artificially infected with a fatal disease in order to provide the serum for an injection to immunise the remaining 99 islanders against the disease or (b) foregoing the immunization programme, in which case there would be a \(10^{-2}\) probability that an epidemic of the disease would kill all 100 of the islanders, so that in both cases the expected number of fatalities is one. The results were as follows:

<table>
<thead>
<tr>
<th></th>
<th>Prefer 1a (Sacrifice one islander)</th>
<th>Prefer 1b ((10^{-2}) chance of all 100 islanders dying)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preference group</td>
<td>13%</td>
<td>87% ((N = 53))</td>
</tr>
<tr>
<td>Fairness group</td>
<td>8%</td>
<td>92% ((N = 53))</td>
</tr>
</tbody>
</table>
Scenario 1', which was also intended to test for ex post equity-proneness, took a similar form to Scenario 1 but in the "no-immunization" option exposed all 100 students to independent probabilities of 10^{-2} of dying. It was explained to subjects that this would mean that there was 36.6% chance that no one would die: a 37% chance that one person would die: an 18.5% chance that two people would die; and so on. As in Scenario 1, in both options the expected number of fatalities is one. The results were as follows:

<table>
<thead>
<tr>
<th>Preference group</th>
<th>Prefer 1a' (Sacrifice one islander)</th>
<th>Prefer 1b' (Independent 10^{-2} chance of death for each islander)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preference group</td>
<td>21%</td>
<td>79% (N = 53)</td>
</tr>
<tr>
<td>Fairness group</td>
<td>9%</td>
<td>91% (N = 53)</td>
</tr>
</tbody>
</table>

Interestingly, these results are at direct variance with the findings of a study reported in Hammerton et al. (1982) which employed a scenario very similar to that of Keller and Sarin's Scenario 1. (The latter was, in fact, a modified version of that in Hammerton et al.) In fact, in the Hammerton et al. study more than 80% of respondents expressed a preference for the "sacrifice one islander" option, though it should be said that in that study the (as yet unidentified) islander had returned from the mainland carrying the disease, so that it may be that in some measure considerations of "retributive justice" drove the Hammerton et al. result.

Scenario 2, which was intended (i) to test for ex ante equity-proneness and (ii) to avoid the possibility that the desire to prevent one certain death might drive the results (as the authors concede might have been the case with scenarios 1 and 1'), involved the choice between (a) exposing all 100 islanders to independent probabilities of 10^{-1} of dying and (b) exposing 50 islanders to independent probabilities of 5 \times 10^{-2} of dying and the remaining 50 to independent probabilities of 1.5 \times 10^{-1} of dying, so that in both options the expected number of fatalities is 10. The results were as follows:

<table>
<thead>
<tr>
<th>Preference group</th>
<th>Prefer 2a (Independent 10^{-1} chance of death for each islander)</th>
<th>Prefer 2b (Independent 5 \times 10^{-2} chance of death for each of 50 and independent 1.5 \times 10^{-1} chance of death for each of remaining 50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preference group</td>
<td>98%</td>
<td>2% (N = 53)</td>
</tr>
<tr>
<td>Fairness group</td>
<td>94%</td>
<td>6% (N = 53)</td>
</tr>
</tbody>
</table>

In each of the scenarios considered so far, at least one of the pairwise options entails the possibility of several deaths. Scenario 3 was therefore designed to test for ex ante equity-proneness in circumstances in which there would be, at most, one death. In particular, this scenario involved the choice between (a) selecting one person (presumably at random) from a group of ten rescuers to take a ten-hour shift in a rescue attempt during which there would be a 10^{-1} probability of a tunnel collapse, resulting in the rescuer's death and (b) having each of the ten rescuers take a one-hour shift, so that each rescuer would face a 10^{-2} probability of death, with the events of death for each of the rescuers being mutually

83
exclusive, so that there would be, at most, one death. In both options the expected number of fatalities is 0.1

The results were as follows:

<table>
<thead>
<tr>
<th></th>
<th>Prefer 3a (One person faces $10^{-4}$ chance of death)</th>
<th>Prefer 3b (Ten people each face mutually exclusive chance of $10^{-2}$ of death)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preference group</td>
<td>17%</td>
<td>83% (N = 53)</td>
</tr>
<tr>
<td>Fairness group</td>
<td>4%</td>
<td>96% (N = 53)</td>
</tr>
</tbody>
</table>

Scenario 4, like Scenario 1 was intended to test for ex post equity-proneness but in a somewhat different context. This scenario involved a situation in which 50 miners were trapped in Location A and a further 50 miners were trapped in Location B. The choice was then between (a) mounting a rescue attempt which, if successful, would save the lives of all 100 miners, the probability of success being 0.5 and (b) focusing the rescue activities on one location, chosen at random with probability 0.5 for each location, in which case all 50 miners at that location would be saved while all 50 miners at the other location would die. In both options the expected number of fatalities is therefore 50. The results were as follows:

<table>
<thead>
<tr>
<th></th>
<th>Prefer 4a (0.5 chance that all 100 miners will die)</th>
<th>Prefer 4b (50 miners die for certain. Group selected at random)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preference group</td>
<td>91%</td>
<td>9% (N = 53)</td>
</tr>
<tr>
<td>Fairness group</td>
<td>91%</td>
<td>9% (N = 53)</td>
</tr>
</tbody>
</table>

In the scenarios considered so far there are no benefits or costs to the individuals involved in the scenarios, other than variations in the risk of death. By contrast, in the final three scenarios additional considerations are introduced. In particular, in Scenario 5 one individual in a community of 100 has chosen to have his house built in a remote (and presumably picturesque) hillside location, knowing that an excessive rainstorm could cause flooding and landslides that would endanger his life. In the event, such a rainstorm has occurred and if he is not rescued the individual will die. The essential choice in this scenario is between (a) having a flood protection team continue to monitor a dam which, if left unattended, would be at risk of collapsing or (b) sending the team to rescue the individual. If option (a) is chosen then the individual in the remote house will die for certain, whereas if option (b) is chosen then there is a $10^{-2}$ probability that the unattended dam will collapse, in which case all 100 residents of the area, including the individual in the remote house, will die. In both options the expected number of fatalities is one.
The results were as follows:

<table>
<thead>
<tr>
<th>Preference group</th>
<th>Prefer 5a (Identified individual dies for certain)</th>
<th>Prefer 5b (10⁻² chance of all 100 individuals dying)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>42%</td>
<td>58%</td>
</tr>
<tr>
<td>(N = 53)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fairness group</td>
<td>47%</td>
<td>53%</td>
</tr>
<tr>
<td>(N = 53)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Clearly, in this case only a marginal majority of respondents displayed *ex post equity-proneness*, in contrast to the substantial majority in Scenario 1, suggesting that at least some respondents took the view that because the individual who would die under option (a) had knowingly assumed a risk in order to enjoy the benefit of living in a preferred location, then this to some degree offset the risk inequity of option (a) vis-à-vis option (b).

Scenario 6 involved two equal-sized communities living on an island. Community 1 enjoys the employment and income generated by an industrial plant located within that community while Community 2 does not. However, the industrial plant also produces toxic waste and the choice is between (a) storing that waste in Community 1, in which case 10 people in that community will die from exposure to the toxic effects of the waste or (b) storing half the waste in Community 1 and half in Community 2, in which case 5 people from each of the two communities will die. The results were as follows:

<table>
<thead>
<tr>
<th>Preference group</th>
<th>Prefer 6a (Store all waste in Community 1)</th>
<th>Prefer 6b (Store half waste in each Community)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>89%</td>
<td>11%</td>
</tr>
<tr>
<td>(N = 53)</td>
<td></td>
<td>(N = 53)</td>
</tr>
<tr>
<td>Fairness group</td>
<td>87%</td>
<td>13%</td>
</tr>
<tr>
<td>(N = 53)</td>
<td></td>
<td>(N = 53)</td>
</tr>
</tbody>
</table>

In this case, it is clear that the employment and income benefits enjoyed by Community 1 have substantially overridden considerations of *ex ante equity-proneness* as far as risk is concerned.

Finally, Scenario 6’ was identical to Scenario 6 save that in this case the industrial plant is located halfway between the two communities so that each benefits equally from the employment and income generated by the plant. In this case the results were as follows:

<table>
<thead>
<tr>
<th>Preference group</th>
<th>Prefer 6a’ (Store all waste in Community 1)</th>
<th>Prefer 6b (Store half waste in each Community)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6%</td>
<td>94%</td>
</tr>
<tr>
<td>(N = 53)</td>
<td></td>
<td>(N = 53)</td>
</tr>
<tr>
<td>Fairness group</td>
<td>6%</td>
<td>94%</td>
</tr>
<tr>
<td>(N = 53)</td>
<td></td>
<td>(N = 53)</td>
</tr>
</tbody>
</table>

In this case, the equal division of the employment and income benefits between the two communities appears to have ensured that *ex ante equity-proneness* re-emerges as the primary consideration in the choice between the two options.

The Maier et al Austrian study employed a questionnaire based on that used by Jones-Lee et al (1985). This study involved a small non-random sample drawn from Vienna and Neulengbach (a rural area close to Vienna). Direct interviews produced ninety-eight completed questionnaires. Again, the results were very similar to those of the Jones-Lee et al study. For example, the value of “own” statistical life for road safety was about A$40,000,000 (presumably in 1988 prices) which, converted to £-Sterling in 1994 prices, becomes £2,830,000. There was also evidence of an “inverted-u” age relationship and of a positive income elasticity for the value of statistical life - though no numerical estimate can be obtained since the regression analysis employed a simple medium/high income dummy.

This study, based on what was arguably a hybrid of the willingness to pay and "matching" approaches, sought to determine the relative priorities that people place on the prevention of a given number of deaths by different causes. Thus, respondents were first asked to consider eight different safety improvements, each of which would prevent ten fatalities. These included reduction in workplace exposure to carcinogenic chemicals; construction of bicycle lanes; regulations concerning the maximum gap between crib slats to reduce the risk of strangulation for young children; the removal of dangerous roadside obstacles (such as trees and boulders); an increase in the number of paramedics and ambulances to deal with heart-attack victims; the enforcement of standards to reduce air pollution; the construction of road crash barriers; and the enforcement of standards to prevent falls by construction workers. Viewed from the perspective of a policy maker, and assuming that there was insufficient funding to carry out all eight programmes, respondents were then asked to prioritize the programmes. Having carried out the prioritization exercise respondents were asked to give a rating of 10 to the programme that they had regarded as having the lowest priority and then to rate the other programmes in terms of the relative values that they placed upon them, again viewed from the perspective of a policy maker. Thus, if they thought that programme X merited twice the expenditure per fatality prevented relative to the expenditure per fatality prevented by the lowest-ranked programme, then programme X was to be given a rating of 20.

The study, which was presumably carried out in the late 1980’s, involved three small convenience samples, the first consisting of 190 undergraduates on an introductory psychology course at San Diego State University; the second comprising 35 students on a health and safety policy course at the University of California, San Diego; and the third a group of retirees who were participating in an Institute of Continued Learning Course at the University of California.

Essentially, Mendeloff and Kaplan found that while the difference between any particular respondent’s maximum and minimum ratings could be quite large, when aggregated across respondents these differences fell substantially, partly as a result of disagreements between respondents about the prioritization of the safety programmes. Thus, while the median difference between each respondent’s maximum and minimum ratings varied between five-fold and seven-fold across the three samples, when one considers the 28 ratio comparisons that are possible given eight safety programmes, the median ratio taken across all respondents within each sample rarely exceeded two-fold and the largest of these median ratios was 3.5 fold, (in the UCSD sample for the prevention of crib slat deaths for young children vs the prevention of delayed deaths caused by air pollution). Mendeloff and Kaplan also examined the impact of framing and the extent of prior discussion of key issues by varying those aspects of the study design across subgroups in the SDSU sample. While variations in these factors did produce some effects, the most dramatic impact resulted from provision (vs non-provision) of information concerning the average age of potential victims of the eight different causes of death and the latency period before the beneficial effects of the safety programme would materialise.

Finally, the authors stress that since respondents were asked to prioritize and rate the eight programmes in the role of policy maker, responses are probably more reflective of the social and ethical considerations that participants felt should underpin such decisions rather than the narrower more self-interested factors, such as aversion to physical risk and perceptions of "own exposure" that are likely to determine responses to more conventional questions concerning willingness to pay for "own" safety.

While the underlying methodology and approach employed in this study parallel those employed in Jones-Lee et al (1985), the survey instrument embodied some important modifications and refinements. Amongst the most significant of these was that in many cases individual marginal rates of substitution of wealth for risk were estimated on the basis of respondent’s subjective perceptions of risk levels, as opposed to the “objective” counterparts to these risk levels employed in the Jones-Lee et al study. These subjective perceptions were elicited essentially by providing respondents with “baseline” risk levels - such as the annual risk of a motorcycle or bus fatality - and then asking for respondents’ estimates of the corresponding risks to car drivers and passengers. A second difference between this study and that conducted by Jones-Lee et al was that questions asking directly about willingness to pay for reductions in the risk of non-fatal injury were also included.

The survey, conducted in Sweden during the period from Autumn 1986 to Spring 1987, was based on a random sample of 1,000 individuals. Respondents were not interviewed directly, but instead were asked to complete a postal questionnaire. There was, however, extensive telephone contact with respondents to resolve misunderstanding and to encourage respondents to complete and return the questionnaire. The response rate to the survey was 50.6 per cent, producing 506 usable questionnaires.

The contingent valuation questions, all of which involved risk reductions expressed as x in 100,000, comprised:

- Q14(b) concerning annual willingness to pay for a 25% reduction in the respondent’s own estimate of annual risk of a fatal heart attack.

- Q17 concerning willingness to pay for a car safety feature which would reduce the risk of a fatal accident for the driver and passengers by various amounts. In particular, Q17(a) involved a 50% reduction in the respondent’s own risk estimate and applied only to the driver, while Q17(b) and Q17(c) involved 25% and 10% reductions in the same risk. Q17(d) then involved a 50% reduction in the respondent’s own risk estimate and applied to the driver and all passengers.

- Q18 concerning willingness to pay to rent a safer car for a single trip where the safer car would afford a 50% reduction in the respondent’s own estimate of the annual risk of a fatal car accident converted to the risk for a single trip and would be applicable only to the driver.

- Q27 concerning willingness to pay additional tax for a 50% reduction in the respondent’s own estimate of the risk of a fatal traffic accident, where the risk reduction would be applicable to all Swedish citizens. Q27(a) involved an additional tax on petrol while Q27(b) concerned a lump sum tax.

In general, the survey appears to have been very successful, judged by various relevant criteria and to have produced results that are, with one or two exceptions, gratifyingly similar to the findings of the study reported in Jones-Lee et al (1985). In particular, the values of “own” statistical life implied by mean responses from the sub-sample of individuals whose subjective perceptions of the risk of being killed as a driver or passenger corresponded to the “objective” level ranged from SEK 14,500,000 (Q27(a)) to SEK 17,600,000 (Q17(a)) in 1986 prices, while those implied by median responses ranged
from SEK 4,000,000 (Q27(a)) to SEK 8,000,000 (Q17(a)). In 1994 prices these convert
to £ Sterling ranges of about £1,910,000 to £2,320,000 and £530,000 to £1,060,000.
Income elasticities of the value of statistical life were also very similar to those found in
studies reported in Blomquist (1979) and Jones-Lee et al (1985) studies. Indeed, the only
major difference between Persson and Cedervall’s findings and those of other
questionnaire-based studies concerns the relationship between individual valuation of
safety and age. For road safety, Persson and Cedervall’s regression results imply that
individual valuation of safety is either monotonically declining with age or follows a
“(non-inverted-)u” age relationship. The “inverted-u” age relationship emerges only in
the case of questions concerning willingness to pay for reduction in the risk of heart
attack.

Finally, Persson and Cedervall’s estimates of willingness to pay based values of avoidance
of non-fatal injuries involving (a) facial lacerations, and (b) concussion were SEK
5,200,000 and SEK 4,200,000 respectively based on mean responses and SEK 1,000,000
and SEK 800,000 based on medians. In 1994 prices these convert to £-Sterling figures of
£690,000; £550,000; £130,000 and £105,000 respectively.

This is the first published account of a study that employs marginal “risk-risk” (RR) questions, as well as CV questions, as a means of estimating WTP based values of safety.

The essence of the marginal risk-risk approach, first proposed in Jones-Lee, Hammerton, and Abbott (1983) and further developed in Jones-Lee (1989), is as follows. Suppose for simplicity that death and non-fatal injury are mutually exclusive events and denote the probabilities of these events for a forthcoming period by $\overline{p}$ and $\overline{q}$ respectively. Now suppose that the individual is indifferent between a reduction $\delta p$ in $\overline{p}$ on the one hand, and a reduction $\delta q$ in $\overline{q}$ on the other. It is then straightforward to show that for an expected utility maximizer - or for a non-expected utility maximizer whose preferences satisfy the property of “betweenness”:  

$$\frac{m_t}{m_D} = \frac{\delta p}{\delta q}$$  

(1)

where $m_t$ and $m_D$ denote the individual’s marginal rates of substitution of wealth for the risks of non-fatal injury and death, respectively.

At a slightly more general level, suppose that the individual is indifferent between reductions $\delta p_1$ in $\overline{p}$ and $\delta q_1$ in $\overline{q}$ on the one hand, and reductions $\delta p_2$ in $\overline{p}$ and $\delta q_2$ in $\overline{q}$ on the other, where $\delta p_1 > \delta p_2$ and $\delta q_1 < \delta q_2$. The counterpart to equation (1) is then

$$\frac{m_t}{m_D} = \frac{\delta p_1 - \delta p_2}{\delta q_2 - \delta q_1}$$  

(2)

In fact, the risk-risk questions employed by Viscusi *et al.* rely on this second result.

The survey involved two questionnaire versions administered to respondents interactively by computer. In both versions the computer programme began by enquiring about the respondent’s personal characteristics and then proceeded to a suite of questions concerning the respondent’s state of health and familiarity with the symptoms of chronic bronchitis (CB). These questions were also designed to provide the respondent with a comprehensive account of the nature of these symptoms.

The computer programme then proceeded to ask the respondent to suppose that he/she faced the choice of moving to one or other of two areas, both of which were safer than the area in which the respondent currently lived in terms of both CB risk and the risk of automobile accidents. Denoting an area in which the annual risks of CB and automobile deaths are $x \times 10^{-5}$ and $y \times 10^{-5}$ respectively by $(x, y)$, respondents were first presented with a “dominant choice” risk-test question involving Area A in which the risks were $(75,$

---

1 Essentially betweenness requires that if an individual is indifferent between two risky prospects A and B then he will also be indifferent between either of these prospects and any probability mixture of them, so that his indifference loci in the Marschak Machina triangle are linear (but, of course, in the case of a non-expected utility maximizer not parallel).
15) and Area B in which the risks were (55, 11). Over 80% of respondents made the correct choice of Area B and of the remaining 20%, only very few failed to reverse their choice after further explanation. The latter were excluded from the remainder of the study.

The next set of questions, common to both versions of the questionnaire, involved risk-risk choices designed to yield ratios of marginal rates of substitution from equation (2). This suite of questions began by presenting the respondent with the choice between Area A with (CB, automobile death) risks of (75, 15) and Area B with risks of (55, 19). If the respondent chose Area B then the computer programme presented the respondent with successive increments in the Area B CB risk until indifference was achieved. If, by contrast, the respondent initially chose Area A then the computer programme presented the respondent with successive decreases in the Area B automobile death risk.

When the risk-risk questions had been completed, the Version A programme then presented the respondent with a series of contingent valuation questions involving choices between pairs of areas that differed in their combinations of cost of living and CB risks. The aim of this series of questions was to identify a situation in which the individual was indifferent between Area A and Area B, where Area A's annual cost of living exceeded that of Area B by $x$ but Area A's annual CB risk was $\delta r$ lower than Area B's risk. The individual's MRS of wealth for CB risk is then given by $\frac{x}{\delta r}$.

The essential difference between Version A and Version B of the questionnaire is that Version B's CV questions - again requiring pairwise choices between areas - involved cost of living/automobile risk trade-offs.

The survey itself (presumably carried out in the late 1980s/early 1990s) involved a representative sample of 389 shoppers recruited from a blue-collar mall in Greensboro, North Carolina by a professional marketing firm. In demographic terms, respondents who answered Version B of the questionnaire were very similar to those who answered Version A.

Several patterns of response to any particular suite of trade-off questions were treated as inconsistent and therefore excluded from the analysis. In the case of the risk-risk questions, for example, inconsistent patterns included those in which the respondent began by choosing, say, Area A and continued to do so even when the automobile death risk in Area B had been reduced to the point at which Area B dominated Area A on both risk dimensions. Similarly, responses indicating indifference between the two Areas for all combinations of CB and automobile death risks were treated as being inconsistent. One particular pattern of responses excluded on grounds of "inconsistency" is, however, somewhat disturbing, namely that in which the respondent began by choosing, say, Area A and reversed this choice only when the automobile death risk in Area B was reduced to the point at which Area B dominated Area A on both risk dimensions. These responses were excluded because it was "impossible to interpolate between the trade-offs implied by the last two questions to obtain an indifference point (because the last question yields no rate of trade-off)." Clearly, however, such a response pattern is consistent with any $\frac{m_i}{m_p}$ ratio smaller than $\frac{1}{20}$ (assuming that the penultimate question in the risk-risk sequence involved a choice between CB/automobile death risks of (75, 15) for Area A and (55, 16) for Area B). Similarly, in the case of the direct CV questions, some excluded responses would have been consistent with very large or very small marginal rates of substitution.
In fact, of the 389 sets of responses to the risk-risk questions, 102 (26.2%) were excluded on grounds of inconsistency and of these, 53 (13.6%) were of the "indeterminate" type referred to in the preceding paragraph. In turn, 54 (27.8%) of the 194 sets of chronic bronchitis CV responses were excluded, 29 (14.9%) of these being indeterminate, while 41, (21.0%) of the 195 sets of automobile death risk CV responses were excluded, 17 (8.7%) of these on grounds of indeterminacy.

The ratio of $\frac{m_I}{m_D}$; the value of preventing a statistical case of chronic bronchitis and the value of statistical life for automobile death risk computed from responses not discarded as inconsistent were as follows, (all monetary values have been converted to £-Sterling, 1994):

<table>
<thead>
<tr>
<th></th>
<th>Based on Means</th>
<th>Based on Medians</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{m_I}{m_D}$</td>
<td>0.68</td>
<td>0.32</td>
</tr>
<tr>
<td>$m_I$ for chronic bronchitis</td>
<td>654,000</td>
<td>338,000</td>
</tr>
<tr>
<td>$m_D$ for automobile death risk</td>
<td>6,060,000</td>
<td>1,690,000</td>
</tr>
</tbody>
</table>
This study was designed to produce estimates of a WTP-based value of statistical life and corresponding values for the prevention of non-fatal injuries, for use in transport project appraisal in New Zealand. The study employed five direct contingent valuation questions, all related to road safety, and one involving a time/safety trade-off in the form of a question concerning speed choice in good and bad weather conditions.

The study involved a stratified random sample survey carried out during 1989/90 as part of the New Zealand Ministry of Transport’s periodic household survey and yielded an effective sample of 655, though this was reduced to 629 when respondents who were unable to answer preliminary “risk-test” questions correctly were omitted. The CV questions, all of which involved reductions of risk expressed as x in 10,000 per annum, comprised:

- Q7 concerning willingness to pay for a reduction in own risk by travelling on a safer toll road
- Q11 concerning willingness to pay for a reduction in risk for self and all other family members (if any) by participating in a road safety training course
- Q14 concerning willingness to pay for a reduction in risk for self and all other family members (if any) by fitting an additional car safety feature
- Q20 concerning willingness to pay for a reduction in risk for self and all other family members (if any) by living in an area with safer roads
- Q22 concerning willingness to pay additional taxes for a reduction in risk for self and all other family members (if any) resulting from roadway and pedestrian facility improvements.

In Q7 the annual risk reduction was given as 6 in 10,000 to 3 in 10,000. In all other CV questions the risk reductions entailed a halving of the respondent’s own estimates of the annual risk of a member of his/her household being killed in a road accident.

In all of the CV questions values of statistical life were computed as the mean of \( m_u + \bar{m}_v \) where \( m_u \) denotes the respondent’s marginal rate of substitution of own wealth for own safety and \( \bar{m}_v \) denotes the average of the respondent’s MRS of own wealth for another household member’s safety, \( m_u \) being derived from responses to Q7 and \( \bar{m}_v \) from responses to Qs 11, 14, 20 and 22. In all cases “inconsistent” responses were omitted.

These included zero protest bids, very large bids and bids entailing that \( \bar{m}_v < 0 \).

With the exception of Q20 (concerning willingness to pay to live in a safer area) more than 50% of responses to the CV questions were omitted on grounds of “inconsistency”. Values of \( m_u \) and \( \bar{m}_v \) estimated from the remaining responses were such that, on average \( \bar{m}_v = 1.16 m_u \).

In the case of speed choice questions, values of statistical life were again computed as the mean of \( m_u + \bar{m}_v \) with \( \bar{m}_v \) estimated on the basis of the marginal rate of substitution of
time for risk of death to an average household member, together with an assumed value of travel time.

Converted to £-Sterling in 1994 prices, values of statistical life estimated from the five CV questions and the speed choice question were as follows:

| Q7 | Safe Toll Road | 800,000 |
| Q11 | Road Safety Course | 570,000 |
| Q14 | Safety Feature | 736,000 |
| Q20 | Safer Area | 745,000 |
| Q22 | Taxes for Road Safety | 915,000 |
| Combined Average From All Useable Responses to All Questions | 754,000 |

Regression analysis and breakdowns indicated that, as in other CV studies, only a very small proportion of the overall variability of individual CV responses could be explained by demographic and other variables (typically less than 10%) with income, area of residence (urban vs rural), family size, age and propensity to engage in financial gambles having the most significant impact, the income elasticity of the value of statistical life being in the region of 0.3 - 0.6.

Finally, while the study did include questions aimed at valuing the prevention of serious non-fatal road injuries relative to the value of preventing a fatality, the injury descriptions covered only the extreme cases of a less severe injury from which recovery would be complete within about a month and a severe head injury, with nothing on the spectrum in between. For the former, respondents indicated that on average the prevention of 30 such injuries was equivalent to preventing a fatality, while not surprisingly, severe head injuries were viewed as being as bad as or worse than death by a majority of respondents. These findings are clearly very much in line with the Standard Gamble results reported in Jones-Lee et al (1995).

The purpose of this study was to estimate the rate at which members of the current generation are willing to trade off current life-saving against life-saving at various different times in the future. As such, the authors believed that their “purpose . . . was to put people in the role of social decision-makers: to ask them to choose between two programs that (generally) benefit persons in different generations but whose costs are borne by the present generation”. However, as the authors themselves concede, to the extent that respondents themselves might expect to benefit directly from current safety improvements, but to do so to a lesser extent (or not at all) in the case of future life-saving, there are question marks against whether or not respondents can be treated as having answered the questions “in the role of social decision-makers”.

The questions posed to respondents took the form of dichotomous-choice variants of open-ended matching questions. Thus, respondents were asked to choose between project A which would prevent 100 deaths from pollution in the coming year and project B which would prevent $X$ deaths from pollution in $T$ years time, given that the costs of the two projects are equal and would be borne now, but that only one project can be funded. $X$ was varied randomly across respondents and $T$ was also varied across respondents, being given values of 5, 10, 25, 50 and 100.

The study comprised three separate surveys, each based on telephone interviews of approximately 12 - 15 minutes duration. The first survey involved 1000 Maryland households; the second 1200 households in the Washington DC metropolitan area; and the third a national random sample of 1000 households.

In analyzing their results, the authors treat a respondent who indicated a preference for the prevention of 100 current deaths over the prevention of $X$ (>100) deaths in $T$ years time as applying a discount factor less than or equal to $100/X$ to each future death prevented vis-a-vis a current death prevented. For an individual who discounts future lives saved at a constant exponential rate it then follows that the discount rate, $\rho$, will be at least equal to the value that satisfies:

$$e^{-\rho T} = \frac{100}{X}$$

On the assumption that $\rho$ is normally distributed across the population, the mean and standard deviation of $\rho$ were then estimated from the dichotomous-choice responses using maximum likelihood techniques and were as follows:

| Mean and Standard Deviation of $\rho$ Assuming Constant Exponential Discounting and $\rho$ Normally Distributed |  |
|---|---|---|---|---|---|---|---|
| Mean $\rho$ | Standard Deviation of $\rho$ |  |  |  |  |  |
| Time Horizon ($T$) |  |  |  |  |  |  |
| 5 years | 0.274 | 0.314 |  |  |  |  |
| 10 years | 0.179 | 0.183 |  |  |  |  |
| 25 years | 0.086 | 0.083 |  |  |  |  |
| 50 years | 0.068 | 0.092 |  |  |  |  |
| 100 years | 0.034 | 0.026 |  |  |  |  |
Finally, the authors regressed individual discount rates on a number of explanatory variables including $T$ and a variety of demographic variables, running two separate regressions, one with the data for $T = 5$ and 10 years pooled and the other with the data for $T = 25, 50$ and 100 years pooled. The results of the regression analysis include the finding that for the shorter time horizons $\rho$ is significantly and quite strongly negatively related to $T$, whereas with the longer horizons the relationship is still significant and negative but somewhat weaker, suggesting that overall the relationship between $\rho$ and $T$ is decreasing and convex. As far as the other variables are concerned only age and race had clear significant effects, with the discount rate being, not surprisingly, positively related to age.

The purpose of this study was to test for the asymmetric treatment of gains and losses and the impact of reference point effects (as predicted by Kahneman and Tversky’s Prospect Theory) in the context of the contingent valuation of automobile safety. As suggested below, however, it is also possible to regard the findings of the study as pertinent to an assessment of various other hypotheses generated within the more conventional Neoclassical framework of expected utility theory.

The study involved two sets of experiments. In the first of these, subjects were first asked a variant of the car safety feature willingness to pay (WTP) question included in the Study reported in Jones-Lee *et al* (1985) involving a reduction in the annual risk of death from $10 \times 10^{-5}$ to $5 \times 10^{-5}$. Subjects were then asked a willingness to accept compensation (WTA) counterpart to this question involving an increase in the annual risk of death from $5 \times 10^{-5}$ to $10 \times 10^{-5}$. The second experiment involved a $2 \times 2$ design to provide four subgroups of respondents with four different questions, namely:

- **Q1** WTP for a gain in safety of a new automobile that reduces the risk of death from $p_s$ to $p_b$ ($p_b < p_s$).
- **Q2** WTA for a reduction in safety of a new automobile that increases the risk of death from $p_b$ to $p_s$.
- **Q3** WTP to avoid a reduction in safety of a new automobile that increases the risk of death from $p_b$ to $p_s$, given that the norm is for other makes of automobile of comparable quality to have risk level $p_b$.
- **Q4** WTA to forego a gain in safety of a new automobile that reduces the risk of death from $p_s$ to $p_b$, given that the norm is for other makes of automobile of comparable quality to have risk level $p_s$.

McDaniels then tests three hypotheses on the basis of the findings of these two sets of experiments. In particular, responses to the first experiment, as well as responses to the first and fourth questions in the second experiment, were used to test the hypothesis that differential treatment of gains and losses will lead to a substantial difference between WTP and WTA responses for a given variation in risk, with WTA significantly exceeding WTP (hypothesis 1). In turn, responses to the first and third questions of the second experiment were used to test the hypothesis that the provision of a reference point at (low) risk level $p_b$ in question 3 - and the effective absence of such a reference point in question 1 - would cause WTP for question 3 to exceed WTP for question 1 by a significant amount (hypothesis 2). Finally, responses to the second and fourth questions in the second experiment were used to test the hypothesis that the provision of a reference point at (high) risk level $p_s$ in question 4 - and the effective absence of such a reference point in question 2 - would cause WTA for question 4 to be less than WTA for question 2 by a significant amount (hypothesis 3). While the WTP and WTA questions in the first experiment asked about the respondent’s household's willingness to pay (or accept) in an open-ended format, in the second experiment respondents were presented with dichotomous choice questions which, in the case of the WTP questions, asked whether or not the respondent’s household would be willing to pay a sum of $700 over and above the basic price of $10,000 for the automobile concerned. In the case of the WTA questions in the second experiment the choice was between accepting or rejecting a $700 discount on the basic price of $10,700.

The first experiment, carried out during April and May, 1986, involved a convenience sample consisting of four groups of adults in Pittsburgh, Pennsylvania. The four groups
included parents of children attending a day care centre, together with workers at the centre; professionals and clerical staff in the office of an economic consultancy; residents of a middle-class neighbourhood; and students in a mid-career public management programme. In the case of the first three groups questionnaires were distributed, completed and returned, while in the case of the mid-career students, questionnaires were completed in class. The survey yielded 55 completed questionnaires. By contrast, the second experiment involved a two-page questionnaire administered to 194 subjects in March 1990 in a location adjacent to an automobile dealership in Bellingham, Washington under the supervision of two individuals who had received training in the administration of contingent valuation surveys.

The mean WTP response in the first experiment was $312.50 in 1986 prices. In £Sterling 1994, this converts to £276. However, in the case of the WTA question there were twelve protest responses including five left blank and seven in which the respondent had written in “zero” (in five of the latter the respondent had indicated in a marginal note that he or she would simply not consider buying a less safe car). If these twelve cases are trimmed then the mean WTP response for the remaining 43 respondents was $308.50 while the mean WTA response was $1107.20 i.e. about 3.25 times as high. In £Sterling 1994, these figures convert to £272 and £977. The difference between these mean responses is significant at p = 0.001.

In the case of the second experiment the results were as follows:

<table>
<thead>
<tr>
<th>Responses to Dichotomous Choice WTP and WTA Questions</th>
<th>Percentage “yes” response</th>
<th>Percentage with value &gt; $700</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. Willing to pay $700 for 20% safety improvement</td>
<td>46%</td>
<td>46%</td>
</tr>
<tr>
<td>Q2. Willing to accept $700 instead of 20% safety improvement</td>
<td>17%</td>
<td>83%</td>
</tr>
<tr>
<td>Q3. Willing to pay $700 for 20% safety improvements, given that alternative vehicles have the improvement</td>
<td>60%</td>
<td>60%</td>
</tr>
<tr>
<td>Q4. Willing to accept $700 instead of 20% safety improvement given that alternative vehicles do not have the improvement</td>
<td>56%</td>
<td>44%</td>
</tr>
</tbody>
</table>

One-tailed tests rejected the null hypothesis of no difference between the proportion of “yes” responses at p=0.01 for hypotheses 1 and 3 and at p=0.1 for hypothesis 2.

While McDaniels treats these findings as evidence strongly supporting the predictions of Prospect Theory in the case of hypotheses 1 and 3 and as providing rather weaker support in the case of hypothesis 2, it is possible to place a somewhat different
interpretation upon them. In particular, expected utility theory predicts a strictly
decreasing, strictly concave valuation function for variations in the probability of death
from its existing level - see for example Jones-Lee (1976) - and this too would suggest
hypothesis 1. By contrast, standard Neoclassical theory would treat the response to Q1 as
a compensating variation for a move from p_a to p_b and the response to Q3 as an
equivalent variation for the move from p_b to p_a so that apart from the effective $700
difference in starting wealth as between the two questions, Neoclassical theory would
predict that the responses to these two questions would be effectively equal (or call this
hypothesis 4). Similarly, Neoclassical theory would treat the response to Q2 as a
compensating variation for a move from p_b to p_a and the response to Q4 as an equivalent
variation for the move from p_a to p_b, so that again ignoring the effective $700
difference in starting wealth as between the two questions, Neoclassical theory would
predict equality between the responses (call this hypothesis 5). Clearly hypothesis 5 is
rejected by McDaniels findings at p=0.01. However, hypothesis 4 cannot be rejected at
the conventional significance level of p=0.05. As such, one might be forgiven for
concluding that McDaniels' findings leave honours effectively even as between Prospect
Theory and Neoclassical Theory.

The purpose of this study was to examine the way in which various perceived characteristics of a hazard (such as voluntariness, control and dread), as well as more traditional socioeconomic variables (such as age and income), influence individual willingness to pay for reductions in the risk posed by the particular hazard.

To this end, respondents were asked about their household’s willingness to pay for a 20% reduction in the risks of death from each of ten different hazards, five of which are relatively “well-defined”, with risks that have been quantified on the basis of historical data, while the remaining five are “less well-defined”, with risks that are not readily quantifiable. The well-defined hazards comprised automobiles, civil aviation, power tools, liquefied natural and petroleum gas and a workplace chemical (vinyl chloride monomer). The less well-defined group consisted of chlorinated water, hazardous waste, nuclear energy, sulphur air pollution and electromagnetic fields. Each CV question gave a description of the hazard concerned and the nature of the risk posed (e.g. the possibility of inducing fatal cancer in the case of electromagnetic fields), together with an indication of the way in which additional safety expenditure might reduce the risk in question (e.g. improvements in air traffic control systems in the case of civil aviation or the rerouting of power cables in the case of electromagnetic fields). However, the questions concerning the well-defined hazards differed from those concerning the less well-defined ones in that while the former gave “base” annual fatality rates, the latter merely referred to a 20% reduction in the potential for causing fatalities.

All CV questions were preceded by a series of instructions and points to bear in mind when answering the questions and suggesting various ways in which the respondent might think about them. The CV questions were then followed by a suite of risk-perception questions asking the respondent to rate each of the ten hazards on eight seven-point Likert scales. The eight characteristics comprised voluntariness, severity, knowledge, control, dread, personal exposure, public exposure and overall risk.

The survey, carried out during April and May 1986, was administered to a convenience sample consisting of four groups of adults in Pittsburgh, Pennsylvania. The four groups included parents of children attending a day care centre together with workers at the centre; professionals and clerical staff in the office of an economic consultancy; residents of a middle-class neighbourhood; and students in a mid-career public management programme. In the case of the first three groups questionnaires were distributed, completed and returned, while in the case of the mid-career students, questionnaires were completed in class. The survey yielded 55 completed questionnaires.

Mean (median) household willingness to pay for a 20% reduction in the number of deaths by each cause during the coming year varied from $385.44 ($100) in the case of automobile deaths to $15.30 ($0) in the case of electromagnetic fields. In £-Sterling 1994, these figures convert to £340 (£88) and £14 (£0). Given that data concerning the dependent variable (household willingness to pay) were truncated at zero, the Tobit limited dependent variable model was used for the regression analysis. Furthermore, a “fixed effects” approach was employed - involving a separate dummy variable for each respondent - to take account of omitted variables such as individual risk aversion, familiarity with the risks etc. Using this model the logarithm of household willingness to pay (rather than estimated marginal rates of substitution) was regressed on the logarithms of the risk perception scores and the socioeconomic variables. Three separate regressions were run, one with the data for all the hazards pooled and one each for the five well-defined hazards and the five less well-defined hazards.
In the pooled analysis, dread, personal exposure, public exposure, age and income were all significant at the 1% level with expected signs, while knowledge and control were significant at 5% but had signs opposite to those expected. In the analysis for the well-defined hazards only personal exposure was significant (at 1%) with the expected sign, while in the analysis for the less well-defined hazards only severity (likelihood of being fatal) and dread were significant (at 1%), again with expected signs.

This study, which effectively replicated Viscusi *et al* (1991) save for the automobile death risk questions, was essentially designed to examine the impact of familiarity with chronic lung disease on responses to marginal risk-risk and direct CV questions.

In particular the survey, using a slightly modified version of Viscusi *et al*'s interactive computer programme, was administered to a sample of 189 respondents from the Washington DC area, all of whom had a relative aged 21 or over who suffered from a chronic respiratory condition. Half of this sample was presented with risk-risk and CV questions framed in terms of Viscusi *et al*'s generic description of chronic bronchitis (Version 1), while the other half answered questions in which the reference to a “case of chronic bronchitis” was altered to “a case of chronic respiratory disease like your relative’s” (Version 2). Having removed those cases categorised as “inconsistent” by precisely the same criteria as employed by Viscusi *et al* for each suite of tradeoff questions, the responses to the risk-risk and direct CV questions were then regressed on various explanatory variables, as were the responses to Viscusi *et al*'s original questions, and the results of the two studies compared, adjusting for respondent’s personal characteristics. Essentially Krupnick and Cropper’s main findings were as follows:

- In the case of the responses to the risk-risk questions, familiarity with the symptoms of chronic bronchitis has no significant effect on either the mean or the variance of responses.

- In the case of the responses to the risk-risk questions, the framing of the questions in terms of “a case of chronic respiratory disease like your relative’s” has no significant effect vis-a-vis the generic framing either on the mean or the variance of responses.

- By contrast, in the case of the direct CV questions, familiarity with the symptoms of chronic bronchitis and framing of the questions in terms of “a case of chronic respiratory disease like your relative’s” both have a significant positive effect on willingness to pay to reduce risk even after correcting for the severity of the relative’s respiratory disease in the case of the latter comparison.

While mean and median responses obscure the impact of differences in respondents’ personal characteristics as between Viscusi *et al*'s sample and Krupnick and Cropper’s, or the impact of differences in the severity of relatives’ respiratory disease as between Krupnick and Cropper’s Version 1 and Version 2 respondents, it is nonetheless instructive to consider the mean and median responses.
These were as follows (all monetary values have been converted to £ Sterling, 1994)

<table>
<thead>
<tr>
<th>The Ratio $\frac{m_1}{m_P}$ and the Value of Preventing a Case of Chronic Bronchitis (£-Sterling, 1994)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Version 1</td>
</tr>
<tr>
<td>Version 2</td>
</tr>
<tr>
<td>$m_1$ Version 1</td>
</tr>
<tr>
<td>$m_1$ Version 2</td>
</tr>
</tbody>
</table>

This study was designed to produce estimates of the value of statistical life and values for the prevention of statistical non-fatal injuries for use in the cost-benefit analysis of road projects in Denmark.

The study employed three direct CV questions concerning willingness to pay for reduction in the risk of a fatal road accident, two of these questions being framed in such a way that safety was a private good, while the third concerned a safety improvement that was essentially a local public good. In addition, the questionnaire contained three CV questions concerning willingness to pay for reduction in the risk of three different severities of non-fatal injury.

The study, carried out in March 1993, involved a sample of 1346 individuals over the age of 18 chosen at random from the Danish Central National Register and yielded 945 completed questionnaires, giving a response rate of 70.2%. Interviews were conducted on a one-to-one basis in respondents’ own homes and took on average 43 minutes to complete. Chi-square tests indicated that the sample did not differ significantly from the Danish population above the age of 18 in terms of age, gender, income and annual mileage travelled by car.

Prior to the questions concerning willingness to pay for a reduction in the risk of a fatal road accident, respondents were told that the objective risk of such an accident in Denmark is 11 in 100,000 per annum and were then asked to give an estimate of their own perceived risk. Approximately 40% of respondents believed that their own risk was lower than the objective risk; 50% believed that it was equal to the objective risk and about 10% believed that it was higher. Respondents were then asked the following CV questions.

• Q7 concerning additional willingness to pay to make a 3000 kilometre bus trip in a bus fitted with more efficient brakes and a stronger body than standard, the resultant reduction in the risk of fatality on the trip being 30 per cent.

• Q13 concerning willingness to pay to have an air bag fitted to the steering wheel or glove compartment of the respondent’s own car, the resultant reduction in the risk of fatality being 20 per cent.

• Q19 concerning willingness to pay additional taxes to finance a local road safety improvement that will reduce the risk of fatal injury by 30 per cent.

In all cases respondents were asked to give personal minimum and maximum willingness to pay figures but were apparently not asked for personal “best estimates”. 
Converted to £ Sterling in 1994 prices, values of statistical life estimated on the basis of respondents' own perceptions of their risk of fatality were as follows:

<table>
<thead>
<tr>
<th>Values of Statistical Life (£-Sterling, 1994)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Based on Means</td>
</tr>
<tr>
<td>Min</td>
</tr>
<tr>
<td>Q7. Bus Safety</td>
</tr>
<tr>
<td>Q13. Airbag</td>
</tr>
<tr>
<td>Q17. Local Road Safety</td>
</tr>
</tbody>
</table>

The non-fatal injury CV questions asked about willingness to pay for 30 per cent reductions in the objective risks of a wrist fracture, shin fracture and an open fracture of the femoral bone. As in the case of Jones-Lee et al 1991 non-fatal injury study, the responses to these questions showed a marked insensitivity to the differences in severity of the injuries concerned and their initial risks.

Finally, Tobit analysis of willingness to pay responses to the airbag questions yielded coefficients that were significant at the 5% level for age, income, household car ownership, annual mileage and own perceived risk, with the estimated coefficient for income entailing an income elasticity of the value of statistical life of approximately 1.1.

This study was essentially a pilot survey carried out prior to a main study designed to produce estimates of WTP-based values of prevention of fatal and non-fatal road injuries for use in the appraisal of proposed road projects in Switzerland. The survey involved two distinct questionnaire versions, each presented to 50% of the sample. In the first version the respondent was asked to focus exclusively on his/her willingness to pay for various reductions in "own risk" of fatal and non-fatal injury in a road accident, ignoring the grief and bereavement of relatives, in the event of such an injury. In the other version the respondent was asked to focus exclusively on his/her willingness to pay to reduce the risk of grief and bereavement that he/she would experience in the event of a relative's death or injury, ignoring the relative's physical and mental suffering. In the case of the non-fatal injuries, the descriptions used were very similar to those employed in the 1991 DoT/TRL Study reported in Jones-Lee et al (1995), though the annual risks of the various severities were substantially larger.

The study, carried out in the Canton of Neuchatel in June 1994, involved a quota sample of 100 (50 for each of the two questionnaire versions) stratified by urban/rural domicile, gender, age and social class. Interviews were conducted on a one-to-one basis in respondents' own homes by ten students from the University of Neuchatel who had received extensive prior training in the aims and methods of the study.

The structure of the first version of the questionnaire was broadly similar to that used in the 1991 DoT/TRL Study in that respondents were first asked about their travelling behaviour and were then presented with a "risk-test" question. This was followed by a ranking and scaling exercise for the six injury severities, including death. Respondents were then given a fairly detailed explanation of the nature of the contingent valuation questions that were to follow, together with a list of instructions about what to consider and not to consider in answering these questions (for example, "consider only your own physical and mental suffering"; "completely disregard the inconvenience your friends and relatives might suffer"; "imagine that you will not have any financial problems in the event of an accident, as insurance will cover all expenses" and so on). The CV questions themselves then asked about willingness to pay for 50% reductions in the risk of death and each of three other non-fatal injuries selected as those that the respondent had ranked second, third and fifth after the severe permanently disabling injury had been deleted from the rankings. In each case respondents were asked to suppose that the risk reduction applied only to the particular injury concerned and that all other injury risks were unaffected. Finally, having answered the suite of four CV questions, respondents were asked about their additional willingness to pay for a further halving of the already reduced risk of death.

The structure of the second version of the questionnaire was similar to that of the first version, but respondents were asked to focus on their own grief, distress and inconvenience in the event of a relative being injured and to disregard the physical and mental suffering of the victim concerned.
The annual risk reductions, all presented to a base equal to the population of the Canton of Neuchatel (160,000) were as follows:

<table>
<thead>
<tr>
<th>No hospitalization</th>
<th>320 in 160,000 to 160 in 160,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>No permanent after effects</td>
<td>800 in 160,000 to 400 in 160,000</td>
</tr>
<tr>
<td>Slight permanent disability</td>
<td>320 in 160,000 to 160 in 160,000</td>
</tr>
<tr>
<td>Substantial and permanent restrictions to work and leisure activities</td>
<td>40 in 160,000 to 20 in 160,000</td>
</tr>
<tr>
<td>Death</td>
<td>24 in 160,000 to 12 in 160,000</td>
</tr>
</tbody>
</table>

Responses to the ranking and scaling questions were much as one might have expected and showed no particular anomalies. As in other studies, severe permanent brain damage, whether for oneself or for a relative, was judged worse than death both in terms of the mean and median rankings. In the case of the risk-test questions, however, the results were, prima facie, particularly disappointing, though this is almost certainly the result of the way in which the question was worded. Thus, respondents were presented with a choice of one of two risk reductions, namely:

A. A reduction in the risk of fatal accidents from 20 per 50,000 to 12 per 50,000

or

B. A reduction in the risk of fatal accidents from 5 per 50,000 to 2 per 50,000.

The responses were: A - 32%, B - 66%, Don't know - 2%.

However, as the authors themselves acknowledge, and respondents' reported reasons for their choices confirm, in many cases those who chose B did so because they believed that this would leave them with a lower final overall risk of death (namely 2 in 50,000), since the question did not make it clear that both risks had to be borne, making the initial risk 25 in 50,000 and the final risk 17 in 50,000 if A were chosen and 22 in 50,000 if B were chosen. In the case of the CV questions themselves, of the 100 respondents, 38 gave zero responses at least once.
The reasons given by respondents for zero bids were as follows:

- Low risk or severity not considered as serious
- Confusion over the idea of "insurance," the role of the State, the collective nature of safety
- Zero WTP for death because there is no suffering
- Zero WTP for the second halving of the risk of death
- Other reasons

While the authors do not report the CV results in the form of values of statistical life or injury prevention, but rather as raw monthly WTP amounts or corresponding amounts per injury avoided, the latter have been converted to values of statistical life or injury prevention in the interests of comparability with the findings of other studies. Converted to £-Sterling in 1994 prices the results were as follows:

<table>
<thead>
<tr>
<th>Injury Category</th>
<th>Victims</th>
<th></th>
<th></th>
<th>Relatives</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
<td>n</td>
<td>Mean</td>
<td>Median</td>
<td>n</td>
</tr>
<tr>
<td>No hospitalization</td>
<td>310,000</td>
<td>110,000</td>
<td>49</td>
<td>530,000</td>
<td>170,000</td>
<td>47</td>
</tr>
<tr>
<td>No permanent after effects</td>
<td>180,000</td>
<td>90,000</td>
<td>39</td>
<td>250,000</td>
<td>160,000</td>
<td>44</td>
</tr>
<tr>
<td>Slight permanent disability</td>
<td>550,000</td>
<td>440,000</td>
<td>14</td>
<td>420,000</td>
<td>110,000</td>
<td>7</td>
</tr>
<tr>
<td>Substantial and permanent restrictions to work and leisure activities</td>
<td>5,570,000</td>
<td>2,520,000</td>
<td>45</td>
<td>7,950,000</td>
<td>4,580,000</td>
<td>42</td>
</tr>
<tr>
<td>Death</td>
<td>10,280,000</td>
<td>4,770,000</td>
<td>48</td>
<td>13,670,000</td>
<td>7,640,000</td>
<td>42</td>
</tr>
</tbody>
</table>

While these results suggest that the responses to the CV questions were highly sensitive to the severities of the injuries, the increase in the values of preventing statistical injuries as one moves from the least to the most severe cases and death is, in fact, due in large part to the substantial decreases in the size of the risk reductions concerned. Thus, while the

---

2 It will be recalled that the list of instructions given to respondents prior to the CV questions included a request to assume that insurance would cover all adverse financial consequences of an accident.

3 This has been done by multiplying the monthly willingness to pay amounts by twelve and then dividing by the risk reduction concerned.
values of preventing statistical injuries increase by a factor of between twenty and forty, the raw monthly CV responses increase by a factor of only about three.

While the CV questions were, in the first instance, asked in an open-ended format, those respondents who were unable to answer were provided with a starting sum and then taken through a process of iterative bidding. Results for these respondents, who constituted 12% of the sample, were markedly lower than for those who responded directly to the open-ended questions. In addition, one or other of two different starting points were used in the iterative bidding process for each injury severity, and clear starting point effects were apparent.

Finally, respondents were asked whether they had succeeded in focusing exclusively on loss of life expectancy, pain and discomfort etc. or whether other factors had also impinged upon their responses. In particular they were asked “when you stated your willingness to pay to avoid various injuries, did you think solely of your human costs or did you also consider other factors such as:

- Medical expenses
- Possible loss of income
- Damage to property
- Your relative’s grief, or as applicable, the victim’s physical and mental suffering
- Other(s). Please specify.”

The results were as follows:

<table>
<thead>
<tr>
<th></th>
<th>Questionnaire for Victim</th>
<th>Questionnaire for Relatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of respondents who</td>
<td>28</td>
<td>35</td>
</tr>
<tr>
<td>considered other factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical expenses</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>Possible loss of income</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Damage to property</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Your relative’s grief</td>
<td>14</td>
<td>-</td>
</tr>
<tr>
<td>Victim’s suffering</td>
<td>-</td>
<td>35</td>
</tr>
<tr>
<td>Other reason</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

As the authors note, it seems unsurprising that several respondents were unable to ignore their relative’s grief and bereavement in their responses to CV questions concerning own risk or their relative’s pain and suffering in their responses to CV questions concerning own grief and bereavement. However, the range of other factors that appear to have had an impact on responses is rather worrying.

This study was designed principally to provide a comparison between WTP-based estimates of the value of statistical life obtained from a given sample of individuals by the two main empirical procedures employed so far in the literature, namely revealed preferences (based on compensating wage differentials) and contingent valuation.

Essentially, the study involved asking a sample of employees of firms in the Montreal area (a) to provide answers to various factual questions concerning their occupations which would form the basis for the compensating wage differential analysis and (b) to answer various willingness to pay (WTP) and willingness to accept compensation (WTA) questions concerning work and car safety.

The survey was conducted between May and July 1990 and the sample was drawn from employees of 13 randomly selected firms in the Montreal area having 100 or more employees. Respondents were selected at random during lunch breaks, the aim being to recruit approximately fifteen respondents from each firm. In the event, 191 questionnaires were completed, reflecting a response rate of 68.7% with respect to those individuals approached.

The questionnaire effectively contained one WTP question and one WTA question for job safety, together with corresponding WTP and WTA questions for car safety. In the case of the job safety questions, respondents were first asked to identify the location of their current job on a ten-step risk ladder, each step being associated with a well-known type of job, running from a low risk job (secretary) on the first step to a high risk job (dynamiter in a mine) on the tenth step. While the annual risks of job-related accidental death were estimated from data provided by the Quebec Workers' Compensation Board, it would appear that these risks were not communicated to respondents.

Having identified their current location on the ladder, respondents were then asked how much of an increase in gross weekly wage they would require in order to work at the same job if the risk of accidental death were instead one step higher on the ladder (Q4A). In turn, respondents were asked how much of their gross weekly wage they would forego to work at the same job if the risk of accidental death were instead one step lower (Q5A). Respondents whose initial risk was at the top of the ladder were asked to suppose that their starting point in the WTA question was one step lower (Q4B), while those whose initial risk was at the bottom of the ladder were asked to suppose that their starting point in the WTP question was one step higher (Q5B).

In the car safety WTP question, no risk ladder was used and respondents were instead asked about their willingness to pay to have air bags installed in a new car that they were purchasing, where the effect of the air bags would be to reduce the respondent's own risk of death in a car accident from 4 in 10,000 per annum to 2 in 10,000. Apparently, no reference was made to the effect of the air bags on other car occupants' safety nor is there any indication of how the lump sum purchase price of the air bags was converted to an effective annual payment. In addition, there is no indication of how the car safety WTA question was framed, though one supposes that this took the form of the discount on the purchase price of a new car that would be required if air bags - normally fitted as a standard feature - were to be omitted.
Converted to £-Sterling in 1994 prices the CV estimates of the value of statistical life (based on mean responses) for work and car safety were as follows:

**CV Estimates of the Value of Statistical Life**  
**Based on Mean Responses (£-Sterling 1994)**

<table>
<thead>
<tr>
<th></th>
<th>VALUE (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1A (WTP, Car Safety)</td>
<td>980,000</td>
</tr>
<tr>
<td>Q1B (WTA, Car Safety)</td>
<td>1,760,000</td>
</tr>
<tr>
<td>Q4A (WTA, Job Safety)</td>
<td>16,370,000</td>
</tr>
<tr>
<td>Q4B (WTA, Job Safety) *</td>
<td></td>
</tr>
<tr>
<td>Q5A (WTP, Job Safety)</td>
<td>14,350,000</td>
</tr>
<tr>
<td>Q5B (WTP, Job Safety)</td>
<td>15,090,000</td>
</tr>
</tbody>
</table>

*Only one respondent answered Q4B.*

This study was designed to produce an estimate of a willingness to pay (WTP)-based value of statistical life for use in the cost-benefit analysis of proposed road projects in France. In contrast to other direct CV studies aimed at obtaining estimates of a WTP-based value of statistical life, this study excluded any reference to probabilistic risk reductions and asked instead about household annual willingness to pay for each of six different numbers of lives to be saved during the coming year. These questions were presented in what is clearly a “bottom up” format, with the number of lives saved in each successive question being increased.

The study was carried out in the early part of 1994, and involved a quota sample of 1000 individuals, stratified by various factors such as age, gender, profession and geographical region. Of the 1000 respondents, 982 gave responses to the CV questions, though no response rate, as such, is reported (presumably many more than 1000 people were approached in the first instance). Interviews were conducted face-to-face and, on average, took about 20 minutes to complete.

Respondents were first asked to estimate the annual number of deaths and non-fatal injuries in road accidents in France, the mean and median responses for deaths being 28,345 and 8,000 respectively, while the mean and median responses for non-fatal injuries were 101,000 and 25,000 respectively. Having answered this question, respondents were then told that the true figures were approximately 10,000 and 200,000 respectively. Significant correlation between estimated deaths and injuries and subsequent WTP responses was found only for the subset of respondents who tended to overestimate deaths and injuries, their median WTP responses exceeding the overall median by some 15%.

Respondents were then presented with the suite of six CV questions which asked about household annual willingness to pay (in the form of additional tax earmarked for safety) for the prevention of 50, 100, 500, 1000, 2000 and 5000 road deaths per annum respectively. These questions were presented in an open-ended format. The value of statistical life, $V$, was then computed as:

\[
V = \frac{\text{Number of households in France} \times \text{WTP}}{\text{Number of lives saved}}
\]

Where WTP denotes the mean or median household willingness to pay.
The results, converted to £ Sterling in 1994 prices were as follows

### WTP Responses and Implied Values of Statistical Life

(£-Sterling, 1994)

<table>
<thead>
<tr>
<th>Lives Saved</th>
<th>WTP Mean</th>
<th>WTP Trimmed Mean*</th>
<th>WTP Median</th>
<th>V Mean**</th>
<th>V Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>36</td>
<td>30</td>
<td>12</td>
<td>15,310,000</td>
<td>5,060,000</td>
</tr>
<tr>
<td>100</td>
<td>53</td>
<td>44</td>
<td>12</td>
<td>11,500,000</td>
<td>2,530,000</td>
</tr>
<tr>
<td>500</td>
<td>75</td>
<td>61</td>
<td>24</td>
<td>3,590,000</td>
<td>1,010,000</td>
</tr>
<tr>
<td>1000</td>
<td>96</td>
<td>72</td>
<td>36</td>
<td>2,070,000</td>
<td>760,000</td>
</tr>
<tr>
<td>2000</td>
<td>117</td>
<td>85</td>
<td>36</td>
<td>1,260,000</td>
<td>380,000</td>
</tr>
<tr>
<td>5000</td>
<td>153</td>
<td>97</td>
<td>48</td>
<td>660,000</td>
<td>200,000</td>
</tr>
</tbody>
</table>

* All responses ≥ FF10,000 (1.8% or responses) trimmed
** Based on untrimmed mean responses

Broadly speaking, responses to the sequence of six CV questions displayed one of four patterns which the authors label “constant WTP/Life”, “gradual increase”, “saturation” and “lump sum”. Essentially, the “constant WTP/life” respondents (39) reported WTP sums that increased more or less in proportion to the number of lives saved in each of the six CV questions. By contrast, while “gradual increase” respondents (368) also reported an increased willingness to pay for larger numbers of lives saved, the increase was less than proportionate, so that for these respondents the value of statistical life fell as the number of lives saved increased. In turn, “saturation” respondents (315) initially increased their reported willingness to pay but reached a point part way through the sequence of CV questions at which a mental budget constraint appeared to bind so that no further increase in willingness to pay emerged in subsequent questions. Finally, “lump sum” respondents (165) apparently reported the same willingness to pay for all six CV questions.

Regression analyses were conducted with three specifications, namely linear, semi-logarithmic and a Box-Cox model. Taken as a whole the results of the regression analysis indicate that the CV responses increased significantly with income, household ownership of a bicycle or motorbike (leading presumably to a higher awareness of risk), the size of the urban area within which the household is located, and the number of persons in the household. Variables with significant negative influence on CV responses were age, the presence of children and the interviewee being other than head of household. As in other CV studies, previous experience of a road accident had no significant influence on CV responses. Finally, R² values range from 0.10 in the linear specification to 0.06 and 0.08 in the semi-logarithmic and Box-Cox cases.

This study was designed to produce estimates of WTP-based values of prevention of non-fatat road injuries for use in the cost-benefit analysis of proposed road projects in Sweden. As well as a question asking about the respondent's willingness to pay for a 50% reduction in his/her own subjective perception of the risk of fatal injury, the questionnaire asked about willingness to pay for various reductions in the risks of three severities of non-fatal injury, namely a serious permanently disabling injury (DIS); a serious temporarily disabling injury (TEM); and a slight injury (SLI). Respondents were provided with brief descriptions of the nature of these injuries and their consequences, such as the extent and nature of hospitalization, implications for work and leisure activities, the extent of pain and discomfort and the duration of disability. The injury descriptions were chosen so as to be broadly representative of the spectrum of different severities of non-fatal road injury in Sweden.

The study, carried out during November and December 1993, involved a sample of 1000 individuals aged between 18 and 74 drawn at random from a national register of postal addresses. Questionnaires were posted to subjects' home addresses and subsequent reminders were sent to those who had not responded after two weeks and five weeks respectively. 447 completed questionnaires were received and, of these, 17 were discarded as having WTP responses that were implausibly high in relation to the respondent's income.

In addition to investigating the way in which the severity and duration of a non-fatal injury would affect willingness to pay for a reduction in risk of sustaining the injury, the study also sought to shed light on the impact of the size of the initial risk of the injury and the magnitude of the risk reduction. The questions presented to respondents therefore varied somewhat between the two subsamples. The question concerning willingness to pay for a reduction in the risk of fatal injury was, however, common to both subsamples. In all cases respondents were provided with information concerning the "actual" average annual risk of each severity of injury, though in the case of the fatal injury they were also asked whether they thought that their own risk was above, below, or equal to the average figure and, if above or below, to give an indication of their perception of the risk.

Apart from variation in the mix of severities of non-fatal injury presented to the two subsamples, the main difference between the questions concerned the "actual" average annual risk of the permanently disabling injury. Thus, while subsample 1 was given the true average figure of 24 in 100,000, the figure presented to subsample 2 was 48 in 100,000.
In fact, the base annual risks, p, and risk reductions, $\Delta p$ for each subsample were as follows:

<table>
<thead>
<tr>
<th></th>
<th>Subsample 1</th>
<th>Subsample 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fatal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p = 8 \text{ in } 100,000$</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>$\Delta p = 50%$</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td><strong>Permanently disabling</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p = 24 \text{ in } 100,000$</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>$p = 48 \text{ in } 100,000$</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>$\Delta p = 50%$</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>$\Delta p = 25%$</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td><strong>Temporarily disabling</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p = 138 \text{ in } 100,000$</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>$\Delta p = 50%$</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>$\Delta p = 25%$</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td><strong>Slight</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p = 667 \text{ in } 100,000$</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>$\Delta p = 50%$</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>$\Delta p = 25%$</td>
<td></td>
<td>√</td>
</tr>
</tbody>
</table>
Converted to £-Sterling in 1994 prices, estimates of the value of statistical life and the value of preventing statistical non-fatal injuries were as follows:

**Subsample 1 : Value of Statistical Life and Values of Preventing Statistical Non-Fatal Injuries (£-Sterling, 1994)**

<table>
<thead>
<tr>
<th></th>
<th>Based on Means</th>
<th>Based on Medians</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fatal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p = 50%</td>
<td>2,750,000</td>
<td>1,090,000</td>
</tr>
<tr>
<td><strong>Permanently Disabling</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p = 50%</td>
<td>1,050,000</td>
<td>580,000</td>
</tr>
<tr>
<td>p = 25%</td>
<td>1,660,000</td>
<td>720,000</td>
</tr>
<tr>
<td><strong>Temporarily Disabling</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p = 50%</td>
<td>120,000</td>
<td>63,000</td>
</tr>
<tr>
<td>p = 25%</td>
<td>200,000</td>
<td>77,000</td>
</tr>
</tbody>
</table>

**Subsample 2 : Value of Statistical Life and Values of Preventing Statistical Non-Fatal Injuries (£-Sterling, 1994)**

<table>
<thead>
<tr>
<th></th>
<th>Based on Means</th>
<th>Based on Medians</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fatal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p = 50%</td>
<td>3,140,000</td>
<td>1,090,000</td>
</tr>
<tr>
<td><strong>Permanently Disabling</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p = 50%</td>
<td>660,000</td>
<td>360,000</td>
</tr>
<tr>
<td>p = 25%</td>
<td>1,100,000</td>
<td>430,000</td>
</tr>
<tr>
<td><strong>Temporarily Disabling</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p = 50%</td>
<td>220,000</td>
<td>63,000</td>
</tr>
<tr>
<td><strong>Slight</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p = 50%</td>
<td>24,000</td>
<td>13,000</td>
</tr>
<tr>
<td>p = 25%</td>
<td>42,000</td>
<td>13,000</td>
</tr>
</tbody>
</table>
Regression analysis of the natural logarithm of CV responses for reduction in the risks of fatal and permanently disabling injuries on the natural logs of a number of variables such as household income, age, number of km travelled by car each year, the respondent's subjective assessment of the risk of a fatal car accident for him or herself and a non-log transformed accident experience dummy indicated that income, km travelled by car and subjective estimates of risk, but not accident experience, tended to have significant positive effects on willingness to pay to reduce risk, with income elasticities of 0.46 and 0.36 for fatal and permanently disabling injuries respectively.

Finally, the authors computed individual ratios of marginal rates of substitution of wealth for the risks of the various severities of injury relative to the corresponding marginal rate of substitution for death. With all zero bids excluded, the means and medians of these individual ratios were as follows:

<table>
<thead>
<tr>
<th>Ratios of Marginal Rates of Substitution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Fatai</td>
</tr>
<tr>
<td>Mean Ratio 1.000</td>
</tr>
<tr>
<td>Median Ratio 1.000</td>
</tr>
<tr>
<td>Permanently Disable</td>
</tr>
<tr>
<td>Mean Ratio 0.637</td>
</tr>
<tr>
<td>Median Ratio 0.333</td>
</tr>
<tr>
<td>Temporarily Disable</td>
</tr>
<tr>
<td>Mean Ratio 0.093</td>
</tr>
<tr>
<td>Median Ratio 0.058</td>
</tr>
<tr>
<td>Slight</td>
</tr>
<tr>
<td>Mean Ratio 0.014</td>
</tr>
<tr>
<td>Median Ratio 0.011</td>
</tr>
</tbody>
</table>

The authors remark that these mean and median ratios are substantially smaller than the corresponding ratios reported in O’Reilly et al (1994) and Jones-Lee et al (1995), and offer various possible explanations for this disparity. However, it would seem that the most obvious and plausible reason lies in the fact that with the exception of the fatal injury, the risk reductions used in this study were very much larger than those employed in the study reported in O’Reilly et al (1994) and Jones-Lee et al (1995) and increased at a very much more substantial rate as one moved from the more to the less severe injuries.

The purpose of this study, which involved a nationally representative sample survey funded by the then U.K. Department of Transport (DoT), was to estimate willingness to pay (WTP)-based values for the prevention of various severities of non-fatal road injury to complement the WTP-based value for the prevention of a road fatality adopted by the DoT in 1988.

Following extensive piloting, carried out during 1990, it was decided to base the main survey on two variants of the questionnaire, one of which contained a suite of contingent valuation questions concerning an annually renewable car safety feature which would reduce the “self-only” risks of death and different severities of non-fatal injuries by various amounts, while the other version of the questionnaire was based on a series of standard-gamble (SG) questions. Essentially, an SG question seeks to determine the probability, \( \pi \), of treatment failure at which an individual would be indifferent between (a) the certainty of a particular severity of non-fatal injury and (b) a treatment for the injury which, if successful, would restore the respondent to his/her normal state of health, but if it failed would result in an outcome that the respondent judged worse than the injury itself. In the majority of SG questions in the survey this worse outcome was specified as death. Denoting the individual’s marginal rate of substitution (MRS) of wealth for the probability of sustaining the non-fatal injury by \( m_I \) and the corresponding MRS for the probability if death by \( m_K \), it can then be shown that under a wide range of circumstances, including standard expected utility maximisation, at least to a good approximation,

\[
\frac{m_I}{m_K} = \pi.
\]

Clearly then, an SG question is intended to provide an estimate of \( m_I \) relative to \( m_K \), rather than an “absolute” estimate of \( m_I \) itself. Thus, in the interest of comparability of the CV and SG results, the former were presented not only in absolute terms but also as means and medians of the individual ratios, \( \frac{m_I}{m_K} \), for the different severities of non-fatal injury.

The main survey was carried out during July, August and September 1991 and was based on a nationally representative sample from Great Britain, stratified by geographical region, social class and car ownership, selected by the Office of Population Censuses and Surveys. CV and SG questionnaire versions were administered on an alternating basis, producing 414 completed CV questionnaires and 409 completed SG questionnaires. The survey was carried out by 42 professional interviewers from the Transport Research Laboratory and interviews were conducted on a face-to-face basis in respondents’ homes. The average time taken to complete the questionnaires averaged 45 minutes for the CV version and 39 minutes for the SG version.

In addition to the CV and SG questions themselves, both versions of the questionnaire involved questions concerning ranking and scaling of the various injury severities, as well as an exercise requiring the prioritisation of various injury risk reductions. A follow-up survey was also carried out on a sub-sample of respondents from the main survey. Taken together with the CV and SG results themselves, the results of these other questions and the follow-up survey provide a rich body of data on which tests of internal consistency, validity and reliability can be conducted. However, these results and the outcome of the tests themselves will not be reported here and the interested reader is referred to Jones-Lee, Loomes, O’Reilly, and Philips (1993) for a full discussion.

While the study did consider slight non-fatal injuries, the principal focus was upon those non-fatal injuries classified as “serious” in U.K. transport statistics. In particular, eight
types of injury spanning the very wide spectrum classified as “serious” were selected and brief “bulletted” injury description cards were prepared for each of these together with a card for normal health and one for death. The full set of cards are reproduced below. In what follows each injury description will be referred to by the code letter shown on the card. In addition, the MRS of wealth for each injury category will be denoted by $m_K$ for death, $m_R$ for injury R etc. while the weighted average of the MRS for serious injuries (with weights equal to the conditional probability of each severity) will be denoted by $m_I$.

The CV questions involved six risk reduction scenarios, each illustrated by a showcard consisting of a grid containing 100,000 squares of which x had been blocked out to convey the idea of “x in 100,000”. Three of these scenarios involved reductions in the annual risks of K, R and S by 4 in 100,000, while the remaining three involved reductions in the annual risks of S, W and X by 12 in 100,000. In what follows the 4 in 100,000 reduction in the risk if S is referred to as S1 while the 12 in 100,000 reduction in the risk of S is referred to as S2.

In turn, the SG question involved eight scenarios, which are all shown on the following page. In the first of these the certain adverse health state was R and the outcome of treatment failure was K. This scenario will be denoted by R vs J/K. Using the same notational convention, the other scenarios were S vs J/K, S vs J/R, X vs J/R, X vs J/K, X vs J/S, W vs J/S and W vs J/X.
## INJURY / HEALTH STATE CARDS

<table>
<thead>
<tr>
<th>J</th>
<th>Your normal state of health</th>
</tr>
</thead>
</table>
| F | 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
| G | 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 your work and leisure activities - possibly some prominent scarring |
| 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 improving
  * after 1-3 years, return to normal health with no permanent disability |
| V | In Hospital
  * no overnight stay in hospital - seen as an outpatient
After Effects
  * moderate to severe pain for 1-4 weeks. Thereafter, 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 |
| S | Immediate unconsciousness, followed shortly by death |

120
In the case of the CV questions, respondents were presented with a vertical list of sums of money running from 0 at the top through to £1, £2, £3, £5, £8, £10 and so on down to £500 and "more than £500" at the bottom. Respondents were then asked to identify amounts they were sure they would be prepared to pay by putting a ∆ alongside each such amount and then to put an X against amounts they felt sure they would not be willing to pay. The aim of this procedure was to identify for each respondent a minimum or personal lower bound (the largest ticked amount) and a maximum or personal upper bound (the smallest crossed amount). Within this range each respondent was asked to place a * against the amount at which he or she would find it most difficult to decide whether or not to buy the safety feature. This was taken as the respondent's "best" estimate. A similar elicitation procedure was used in the SG questions, though in this case the list of sums of money was replaced by a list of chances of failure/success of the risky treatment.

Denoting the MRS of wealth for risk of S based on the 4 in 100,000 reduction in the latter by $m_{s1}$ and that based on the 12 in 100,000 by $m_{s2}$, with the corresponding weighted averages for all severities of serious non-fatal injury denoted by $m_{11}$ and $m_{12}$, the marginal rates of substitution estimated from the CV responses were as follows:

| Contingent Valuation Estimates of Marginal Rates of Substitution (£-Sterling x $10^4$, 1994; n = 414*) |
|---|---|---|---|---|---|---|
| Means (£ x $10^4$) | Medians (£ x $10^4$) |
| best | min | max | best | min | max | best = zero* |
| $m_K$ | 4.56 | 3.70 | 5.82 | 2.70 | 2.03 | 4.05 | 4.4% |
| $m_R$ | 3.79 | 3.10 | 4.88 | 2.03 | 1.35 | 2.70 | 3.7% |
| $m_{s1}$ | 2.76 | 2.18 | 3.67 | 1.35 | 1.08 | 1.62 | 5.6% |
| $m_{s2}$ | 1.11 | 0.89 | 1.50 | 0.68 | 0.45 | 0.90 | 3.8% |
| $m_X$ | 0.91 | 0.71 | 1.25 | 0.45 | 0.36 | 0.54 | 5.0% |
| $m_W$ | 0.82 | 0.65 | 1.10 | 0.36 | 0.27 | 0.54 | 7.4% |
| $m_{11}$ | 1.87 | 1.49 | 2.47 | 0.94 | 0.70 | 1.21 | -- |
| $m_{12}$ | 1.47 | 1.18 | 1.94 | 0.78 | 0.55 | 1.04 | -- |

* Missing cases (predominantly "unable to answer") vary from 12 to 26 but are typically about 20.

*: Computed as a percentage of all non-missing cases (i.e., all cases minus missing cases).
In turn, the means and medians of individual ratios of marginal rates of substitution implied by the CV responses were:

<table>
<thead>
<tr>
<th></th>
<th>Means</th>
<th>Medians</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>best</td>
<td>min</td>
</tr>
<tr>
<td>( \frac{m_x}{m_K} )</td>
<td>0.875</td>
<td>0.675</td>
</tr>
<tr>
<td>( \frac{m_{\delta X}}{m_K} )</td>
<td>0.640</td>
<td>0.486</td>
</tr>
<tr>
<td>( \frac{m_{\delta Y}}{m_K} )</td>
<td>0.262</td>
<td>0.202</td>
</tr>
<tr>
<td>( \frac{m_{\delta Z}}{m_K} )</td>
<td>0.232</td>
<td>0.178</td>
</tr>
<tr>
<td>( \frac{m_{\delta}}{m_K} )</td>
<td>0.210</td>
<td>0.163</td>
</tr>
<tr>
<td>( \frac{m_{\delta}}{m_K} )</td>
<td>0.443</td>
<td>0.344</td>
</tr>
<tr>
<td>( \frac{m_z}{m_K} )</td>
<td>0.352</td>
<td>0.276</td>
</tr>
</tbody>
</table>

* All ratios have been computed with reference to the individual 'best' estimates of \( m_k \).
  All cases where individual \( m_k = 0 \) have been omitted.
  Missing cases vary from 34 to 46, of which 17 are cases in which \( m_k = 0 \)
Finally, the means and medians of individual ratios of marginal rates of substitution implied by the SG responses were:

<table>
<thead>
<tr>
<th>Standard Gamble Results (n = 409*)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Means</th>
<th>Medians</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>best</td>
<td>min</td>
</tr>
<tr>
<td>( \frac{m_r}{m_k} )</td>
<td>0.233</td>
<td>0.203</td>
</tr>
<tr>
<td>( \frac{m_r}{m_k} )</td>
<td>0.151</td>
<td>0.122</td>
</tr>
<tr>
<td>( \frac{m_r}{m_k} )</td>
<td>0.055</td>
<td>0.051</td>
</tr>
<tr>
<td>( \frac{m_r}{m_k} )</td>
<td>0.020</td>
<td>0.022</td>
</tr>
<tr>
<td>( \frac{m_r}{m_k} )</td>
<td>0.117</td>
<td>0.104</td>
</tr>
</tbody>
</table>

* Missing cases (predominantly “unable to answer”) vary from 4 to 29 but are typically between 10 and 20.

: Computed as a percentage of all non-missing cases (i.e., all cases minus missing cases).
\( \times \) \( m_w/m_k \) computed as the product \( m_y/m_k \times m_w/m_S \) has 29 missing cases at “best” but only 5 at “min”

This study, which was commissioned and funded by London Underground Limited (LUL), was intended to provide estimates of the WTP-based value of preventing Underground fatalities relative to the corresponding value for the roads and to explore the reasons why the former might stand at a premium (or discount) in relation to the latter.

Following a pilot study, the main survey was structured around thirty “focus group” meetings, each involving between six and eight members of the public who were both Underground users and car drivers or passengers. The structure of the focus group meetings was as follows:

- Participants were first presented with a number of short statements concerning safety in general and invited to consider these statements in an open-ended discussion, guided by the group moderator (Graham Loomes), and to explore the extent of their agreement or disagreement with these statements.

- Participants were then presented with four statements concerning large-scale as opposed to small-scale Underground accidents and asked to indicate (on an individual, rather than group basis) the extent of their agreement or disagreement with these statements. The statements concerned were based on typical views expressed in the pilot study focus group discussions.

- Following completion of this block of four “agree/disagree” questions, participants were then presented with quantitative questions - again to be answered on an individual, rather than group basis - designed to elicit “scale” premia for the prevention of large-scale Underground accidents relative to small-scale Underground accidents.

- Participants were then presented with four “agree/disagree” questions related to statements concerning small-scale Underground accidents as opposed to small-scale road accidents.

- Finally, participants were presented with quantitative questions designed to elicit “context” premia for the prevention of small-scale Underground accidents relative to small-scale road accidents.

Essentially, the quantitative “scale” questions sought to determine the number of single-fatality Underground accidents on escalators, lifts and platforms that respondents would require to be prevented by a safety programme if the latter were to be judged “equally as good as” a programme of the same cost that would prevent one large-scale Underground accident (such as occurred at King's Cross in 1987) during the same period. In turn, in order to estimate the quantitative “context” premium respondents were asked about the number of single-fatality road accidents that they would require a road safety programme to prevent in the London area if that programme were to be judged “equally as good as” a programme of the same cost that would prevent 25-30 single-fatality Underground accidents during the same period.
Responses to the "agree/disagree" questions related to scale effects were as follows

A. 25-30 deaths in a single Underground accident is worse than 25-30 deaths in separate Underground accidents.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>24%</td>
<td>43%</td>
<td>13%</td>
<td>17%</td>
<td>3%</td>
</tr>
</tbody>
</table>

\( n = 222 \)

B. The experts probably have a pretty good idea about the causes of small-scale Underground accidents and the number of people likely to be killed in them over the next 25 years, but they have much less idea about the chances of large-scale Underground accidents and the number of people likely to be killed in those kinds of accidents over the next 25 years.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>2%</td>
<td>19%</td>
<td>15%</td>
<td>59%</td>
<td>5%</td>
</tr>
</tbody>
</table>

\( n = 221 \)

C. Victims of large-scale Underground accidents are typically not to blame for them, whereas small-scale Underground accidents are likely to be at least partly the victim's own fault.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td>29%</td>
<td>21%</td>
<td>42%</td>
<td>3%</td>
</tr>
</tbody>
</table>

\( n = 218 \)

D. You can be pretty sure that if you spend more money trying to prevent small-scale accidents then fewer will occur; whereas, however much you spend trying to prevent large-scale accidents, it probably won't make much difference: in some cases, they might not happen anyway, and in other cases, some freak set of circumstances may still cause them to occur.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td>25%</td>
<td>15%</td>
<td>50%</td>
<td>5%</td>
</tr>
</tbody>
</table>

\( n = 220 \)
In turn, responses to the “agree/disagree” questions related to context effects were as follows:

E. The thought of being killed in an Underground accident is worse than the thought of being killed in a road accident

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>11%</td>
<td>29%</td>
<td>30%</td>
<td>21%</td>
<td>9%</td>
</tr>
</tbody>
</table>

n = 223

F. London Underground is providing a public service and therefore has a duty to protect the public’s safety, whereas road users have a greater responsibility to look after their own safety.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>2%</td>
<td>6%</td>
<td>5%</td>
<td>58%</td>
<td>29%</td>
</tr>
</tbody>
</table>

n = 220

G. When I travel by Underground, it’s usually because there’s no real alternative, and I therefore have little choice but to accept whatever risks are involved; whereas when I travel by car, it’s usually because I choose to do so, and I therefore voluntarily accept whatever risks are involved.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>2%</td>
<td>19%</td>
<td>11%</td>
<td>56%</td>
<td>12%</td>
</tr>
</tbody>
</table>

n = 224

H. When I’m travelling by car, I have much more control over my own safety than I do when I’m travelling on the Underground.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>2%</td>
<td>17%</td>
<td>13%</td>
<td>51%</td>
<td>17%</td>
</tr>
</tbody>
</table>

n = 223

Finally, responses to the quantitative questions yielded a mean scale premium of -2% (with a standard error of 6%) and a mean context premium of 51% (with a standard error of 7%).

The purpose of this study was to examine the feasibility, reliability and validity of four value-elicitation techniques that might be used to estimate willingness to pay-based monetary values for food safety. The four techniques were:

- Contingent Valuation (CV) questions
- Standard Gamble (SG) questions
- Marginal Risk-Risk (RR) questions
- Matching (or equivalence) questions.

The study, carried out during February, March and April, 1995, was conducted in three stages. In the first stage eleven focus group meetings were held in Newcastle and York. Each focus group comprised between 5 and 8 participants recruited from various local organisations in Newcastle (e.g. Associations for Single Parents and Townswomen’s Guilds) and the parents and other relatives of pupils at a large local comprehensive school in York. The total number of participants was 73. The focus group meetings, which had a semi-structured format, were intended to encourage participants to give thought to issues of safety in general, and food safety in particular, and to introduce the four adverse health states that might result from food poisoning in eggs and which would form the basis of the questions in the individual interviews that would follow in the second stage of the study. Focus group participants were also given the average annual probabilities of contracting each of the four adverse health states.

The second stage of the study then consisted of face-to-face individual interviews, conducted by two members of the research team (Judith Covey and Angela Robinson), in which respondents were presented with one of four variants of a questionnaire containing CV, SG, RR and matching questions.
These four variants reflected a 2 x 2 design in which half of the sample were presented with questions concerning the following adverse health states and associated annual probabilities (all expressed to a base of 6,000,000):

**K**

Stomach pains, diarrhoea, vomiting and fever requiring admission to hospital.
Rapid deterioration of your condition followed shortly by death.

2 in 6,000,000

**B**

Stomach pains, diarrhoea, vomiting and fever sufficiently severe to be admitted to hospital. Following discharge from hospital, your symptoms will gradually improve over the next 15-30 days BUT you will experience periods of stomach discomfort and diarrhoea on average several times a week for the rest of your life.

240 in 6,000,000

**L**

Mild stomach pains, diarrhoea, vomiting and fever. Return to normal health in 2-5 days.

12,000 in 6,000,000
Participants in the other half of the sample were also presented with questions concerning K and L but the following adverse health state was substituted for B:

<table>
<thead>
<tr>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stomach pains, diarrhoea, vomiting and fever sufficiently severe to be admitted to hospital. Return to normal health in 15-30 days, although a limited number of cases can last for a number of months.</td>
</tr>
</tbody>
</table>

240 in 6,000,000

The second aspect of the 2 x 2 design was a 50/50 split between (a) respondents who were presented with a “bottom up” (BU) sequence of CV questions which began with a series of questions about separate 50% reductions in the risks of each of the three adverse health states that might result from food poisoning in eggs and ended with a question in which all three risks were halved simultaneously and (b) respondents who were presented with a “top down” (TD) sequence of CV questions which began with a question in which all risks were simultaneously reduced by 50% and was then followed by a series of questions in which successively smaller subsets of these risks were halved.

In summary, the 2 x 2 design yielded the following four sets of CV questions:

**Version A**  Bottom up sequence of CV questions concerning K, B and L

**Version B**  Top down sequence of CV questions concerning K, B and L

**Version C**  Bottom up sequence of CV questions concerning K, E and L

**Version D**  Top down sequence of CV questions concerning K, E and L
The actual sequences of the CV questions in the four versions of the questionnaire were as follows:

<table>
<thead>
<tr>
<th>Group</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Annual K</td>
<td>Box K</td>
<td>Box B</td>
<td>Box L</td>
<td>Box (K+B+L)</td>
</tr>
<tr>
<td>B</td>
<td>Annual (K+B+L)</td>
<td>Box (K+B+L)</td>
<td>Box (K+B)</td>
<td>Box K</td>
<td>Box L</td>
</tr>
<tr>
<td>C</td>
<td>Annual K</td>
<td>Box K</td>
<td>Box E</td>
<td>Box L</td>
<td>Box (K+E+L)</td>
</tr>
<tr>
<td>D</td>
<td>Annual (K+E+L)</td>
<td>Box (K+E+L)</td>
<td>Box (K+E)</td>
<td>Box X</td>
<td>Box L</td>
</tr>
</tbody>
</table>

Where Box K, Box B, Box E and Box L represent the additional willingness to pay per box for half a dozen of an alternative brand of eggs in which the risk of K, B, E and L respectively are halved. A combination of conditions in brackets indicates that the risks of two or more conditions are reduced simultaneously i.e. Box (K+B+L) represents the additional willingness to pay per box for the alternative eggs in which the risks of conditions K, B and L are all halved. The same interpretation may be put on Annual K etc. Although of course this represents the additional willingness to pay (WTP) per year.

In all four versions of the questionnaire the suite of CV questions was then followed by two SG questions, two RR questions and two Matching questions, in that order. The format of the SG and matching question was similar to that reported above in the summaries of Jones-Lee et al (1995) and Jones-Lee and Loomes (1995). In the case of the RR questions, these differed somewhat from those employed in Viscusi et al (1991) and Krupnick and Cropper (1992). In particular, the RR questions employed in this study sought to determine the largest increase in the risk of, say, condition B that the respondent would be prepared to tolerate in order to eliminate completely the risk of K.

The third and final stage of the study then involved re-convining the focus groups (not necessarily with the same composition as in the first stage) in order to feed back to respondents a simplified summary of the results of the second stage and to provide them with an opportunity to comment on these results and the design of the questionnaire.

Not surprisingly, the findings of the three stages of the study constitute a rich and fairly complex body of data which the authors analyse in some depth. Here we will simply summarise the main quantitative results of the stage 2 individual interviews.

In the case of CV questions the results were as follows:

### Willingness to Pay Responses of Groups A and C

<table>
<thead>
<tr>
<th></th>
<th>Group A (condition B)</th>
<th>Group C (condition E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>18</td>
<td>17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question</th>
<th>Group A</th>
<th>Group C</th>
<th>Group A</th>
<th>Group C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>Q1 Eggs consumed per week</td>
<td>3.4</td>
<td>3.0</td>
<td>3.8</td>
<td>2.0</td>
</tr>
<tr>
<td>Q2 Annual K in £'s</td>
<td>7.90</td>
<td>4.38</td>
<td>5.63</td>
<td>4.00</td>
</tr>
<tr>
<td>Q3 Per box K in pence</td>
<td>35.3</td>
<td>22.5</td>
<td>25.7</td>
<td>25.0</td>
</tr>
<tr>
<td>Implied Annual K in £'s</td>
<td>9.23</td>
<td>4.25</td>
<td>7.11</td>
<td>4.33</td>
</tr>
<tr>
<td>Q4 Per box B or E in pence</td>
<td>37.3</td>
<td>27.5</td>
<td>24.8</td>
<td>15.0</td>
</tr>
<tr>
<td>Q5 Per box L in pence</td>
<td>20.1</td>
<td>15.0</td>
<td>16.5</td>
<td>11.0</td>
</tr>
<tr>
<td>Q6 Per box K+L+(B or E) in pence</td>
<td>45.9</td>
<td>35.0</td>
<td>24.9</td>
<td>25.0</td>
</tr>
<tr>
<td>Sum = Q3 + Q4 + Q5</td>
<td>92.8</td>
<td>61.0</td>
<td>66.9</td>
<td>52.5</td>
</tr>
</tbody>
</table>
Willingness to Pay Responses of Groups B and D

<table>
<thead>
<tr>
<th>Question</th>
<th>Group B (condition B)</th>
<th>Group D (condition E)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>Q1</td>
<td>2.9</td>
<td>2.0</td>
</tr>
<tr>
<td>Q2</td>
<td>10.12</td>
<td>6.25</td>
</tr>
<tr>
<td>Q3</td>
<td>41.7</td>
<td>38.8</td>
</tr>
<tr>
<td>Implied Annual K+L+(B or E) in £'s</td>
<td>11.38</td>
<td>5.25</td>
</tr>
<tr>
<td>Q4</td>
<td>38.0</td>
<td>21.3</td>
</tr>
<tr>
<td>Q5</td>
<td>13.4</td>
<td>0.0</td>
</tr>
<tr>
<td>Q6</td>
<td>17.7</td>
<td>3.0</td>
</tr>
<tr>
<td>Sum = Q4 + Q6</td>
<td>55.7</td>
<td>50.0</td>
</tr>
<tr>
<td>Implied B or E in pence (Q4-Q5)</td>
<td>24.5</td>
<td>17.5</td>
</tr>
<tr>
<td>Implied L in pence (Q3-Q4)</td>
<td>3.7</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Clearly, the willingness to pay figures can be used to estimate marginal rates of substitution of wealth for risk of the adverse health state concerned simply by dividing the implied annual willingness to pay by the appropriate annual risk reduction. In the case of K (the fatal condition), for example, the risk reduction was 1 in 6,000,000 so that estimates of the value of statistical life implied by the Q3 willingness to pay responses of each of the four groups were as follows:

Values of Statistical Life (£-Sterling, 1995)

<table>
<thead>
<tr>
<th>Group</th>
<th>Based on Means</th>
<th>Based on Medians</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>55,380,000</td>
<td>25,500,000</td>
</tr>
<tr>
<td>Group B</td>
<td>25,620,000</td>
<td>0</td>
</tr>
<tr>
<td>Group C</td>
<td>42,660,000</td>
<td>25,980,000</td>
</tr>
<tr>
<td>Group D</td>
<td>71,700,000</td>
<td>0</td>
</tr>
</tbody>
</table>
In turn, the ratios of marginal rates of substitution implied by responses to the CV questions were:

<table>
<thead>
<tr>
<th>CV Ratios of Marginal Rates of Substitution</th>
<th>Groups A and C</th>
<th>Groups B and D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>(m_p/m_k)</td>
<td>0.0086</td>
<td>0.0083</td>
</tr>
<tr>
<td>(m_e/m_k)</td>
<td>0.0078</td>
<td>0.0083</td>
</tr>
<tr>
<td>(m_i/m_k)</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

In the case of the SG questions, which also yield estimates of ratios of marginal rates of substitution, the results were:

<table>
<thead>
<tr>
<th>SG Ratios of Marginal Rates of Substitution</th>
<th>Groups A and B (N=36)</th>
<th>Groups C and D (N=33)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>(m_p/m_k)</td>
<td>0.0486</td>
<td>0.0002</td>
</tr>
<tr>
<td>(m_e/m_k)</td>
<td>0.0365</td>
<td>0.0000</td>
</tr>
<tr>
<td>(m_i/m_k)</td>
<td>0.0006</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

In turn, estimates of ratios of marginal rates of substitution based on responses to the RR questions were as follows:

<table>
<thead>
<tr>
<th>RR Ratios of Marginal Rates of Substitution</th>
<th>Groups A and B (N=36)</th>
<th>Groups C and D (N=33)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>(m_p/m_k)</td>
<td>0.1741</td>
<td>0.0583</td>
</tr>
<tr>
<td>(m_e/m_k)</td>
<td>0.0426</td>
<td>0.0008</td>
</tr>
</tbody>
</table>

Finally, denoting the marginal rate of substitution of wealth for risk of death in a road accident by \(m_p\), heart disease by \(m_e\), cancer by \(m_c\) and, as above, food poisoning by \(m_k\),
estimates of ratios of these marginal rates of substitution based on responses to the matching questions were as follows:

<table>
<thead>
<tr>
<th></th>
<th>Arithmetic Mean</th>
<th>Median</th>
<th>Geometric Mean</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{R}/m_{B}$</td>
<td>0.989</td>
<td>0.875</td>
<td>0.686</td>
<td>69</td>
</tr>
<tr>
<td>$m_{R}/m_{D}$</td>
<td>1.055</td>
<td>0.875</td>
<td>0.688</td>
<td>35</td>
</tr>
<tr>
<td>$m_{V}/m_{C}$</td>
<td>1.199</td>
<td>1.000</td>
<td>0.736</td>
<td>34</td>
</tr>
</tbody>
</table>
## Summary of Contingent Valuation Estimates of the Value of Statistical Life

<table>
<thead>
<tr>
<th>Author</th>
<th>Context</th>
<th>Risk Reduction</th>
<th>Value of Statistical Life (£ Sterling, 1994)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1] Acton</td>
<td>Heart attack</td>
<td>$2 \times 10^5$ p.a.</td>
<td>Based on Means: 69,800 (80,500) (112,000)</td>
</tr>
<tr>
<td>(1973)</td>
<td>U.S.A.</td>
<td></td>
<td>Based on Medians: -</td>
</tr>
<tr>
<td></td>
<td>U.K.</td>
<td></td>
<td>Based on Medians: 132,900 (153,900) (215,800)</td>
</tr>
<tr>
<td>[3] Jones-Lee (1976)</td>
<td>Aviation</td>
<td>$X \times 2 \times 10^6$ per flight</td>
<td>Based on Means: 10,100,000 (11,400,000) (15,100,000)</td>
</tr>
<tr>
<td></td>
<td>U.K.</td>
<td></td>
<td>Based on Medians: 8,590,000 (9,700,000) (12,800,000)</td>
</tr>
<tr>
<td>[4] Maclean</td>
<td>Domestic fire</td>
<td>--</td>
<td>Based on Means: 3,560,000 (3,945,000) (5,007,000)</td>
</tr>
<tr>
<td>(1979)</td>
<td>U.K.</td>
<td></td>
<td>Based on Medians: -</td>
</tr>
<tr>
<td>[5] Frankel</td>
<td>Aviation</td>
<td>$1.5 \times 10^6$ per flight</td>
<td>Based on Means: 19,000,000 (20,500,000) (24,500,000)</td>
</tr>
<tr>
<td>(1979)</td>
<td>U.S.A.</td>
<td></td>
<td>Based on Medians: 4,800,000 (5,180,000) (6,180,000)</td>
</tr>
<tr>
<td>[5] Frankel</td>
<td>Aviation</td>
<td>$1 \times 10^3$ per flight</td>
<td>Based on Means: - 79,800 (86,100) (102,800)</td>
</tr>
<tr>
<td>(1979)</td>
<td>U.S.A.</td>
<td></td>
<td>Based on Medians: -</td>
</tr>
<tr>
<td></td>
<td>U.K.</td>
<td></td>
<td>Based on Medians: 1,330,000 (1,440,000) (1,750,000)</td>
</tr>
<tr>
<td></td>
<td>U.S.A.</td>
<td></td>
<td>Based on Medians: -</td>
</tr>
<tr>
<td></td>
<td>Austria</td>
<td></td>
<td>Based on Medians: -</td>
</tr>
</tbody>
</table>

*In some cases individual studies produced a variety of estimates. These have typically been distilled into a single "representative" figure either by taking the particular author's preferred estimate or by a straightforward process of averaging. However, in some cases individual studies have produced estimates that differ by more than an order of magnitude and these estimates are reported in full.

ψ In each case, the first figure shown has been updated only for inflation. The figures in brackets then reflect further updating for increases in real income per capita, assuming income elasticities of the value of statistical life of 0.3 and 1.0 respectively.
### Summary of Contingent Valuation Estimates of the Value of Statistical Life (K) (Sterling, 1994)

<table>
<thead>
<tr>
<th>Author</th>
<th>Context</th>
<th>Risk Reduction</th>
<th>Value of Statistical Life (£ Sterling, 1994)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Based on Means</td>
</tr>
<tr>
<td>[14] Persson et al (1991)</td>
<td>Roads Sweden</td>
<td>$X \times 10^{-4}$ p.a.</td>
<td>2,120,000 (2,190,000)</td>
</tr>
<tr>
<td>[15] Viscusi et al (1991)</td>
<td>Roads U.S.A.</td>
<td>$X \times 10^{-4}$ p.a.</td>
<td>6,060,000 (6,180,000)</td>
</tr>
<tr>
<td>[16] Miller et al (1991)</td>
<td>Roads New Zealand</td>
<td>$X \times 10^{-4}$ p.a.</td>
<td>800,000 (808,000)</td>
</tr>
<tr>
<td>[21] Kidholm (1995)</td>
<td>Roads Denmark</td>
<td>$X \times 10^{-4}$ p.a.</td>
<td>7,400,000 (7,440,000)</td>
</tr>
<tr>
<td>[22] Schwab - Christe (1995)</td>
<td>Roads Switzerland</td>
<td>$X \times 6.25 \times 10^{-6}$ p.a.</td>
<td>10,280,000 (10,280,000)</td>
</tr>
<tr>
<td>[23] Lanoie et al (1995)</td>
<td>Roads Canada</td>
<td>$X \times 10^{-8}$ p.a.</td>
<td>980,000 (1,013,000)</td>
</tr>
<tr>
<td>[23] Lanoie et al (1995)</td>
<td>Workplace Canada</td>
<td>-</td>
<td>14,700,000 (15,200,000)</td>
</tr>
<tr>
<td>[24] Desaigues et al (1995)</td>
<td>Roads France</td>
<td>50 deaths p.a.</td>
<td>15,300,000 (15,300,000)</td>
</tr>
<tr>
<td>[24] Desaigues et al (1995)</td>
<td>Roads France</td>
<td>500 deaths p.a.</td>
<td>3,590,000 (3,590,000)</td>
</tr>
<tr>
<td>[25] Persson et al (1995)</td>
<td>Roads Sweden</td>
<td>$4 \times 10^{-7}$ p.a.</td>
<td>2,950,000 (2,960,000)</td>
</tr>
<tr>
<td>[26] Jones-Lee et al (1995)</td>
<td>Roads U.K.</td>
<td>$4 \times 10^{-7}$ p.a.</td>
<td>4,560,000 (4,680,000)</td>
</tr>
<tr>
<td>[28] Covey et al (1995)</td>
<td>Food U.K.</td>
<td>$1.67 \times 10^{-7}$ p.a.</td>
<td>48,840,000 (48,840,000)</td>
</tr>
</tbody>
</table>

**These figures are in early 1995 prices.**
MAIL ORDER
HSE priced and free publications are available from:
HSE Books
PO Box 1999
Sudbury
Suffolk CO10 6FS
Tel: 01787 881165
Fax: 01787 313995

RETAIL
HSE priced publications are available from good booksellers

HEALTH AND SAFETY ENQUIRIES
HSE InfoLine
Tel: 0541 545500
or write to:
HSE Information Centre
Broad Lane
Sheffield S3 7HQ

HSE home page on the World Wide Web:
http://www.open.gov.uk/hse/hsehome.htm

CRR 171

£37.50 net