


IN MEMORIAM

Michael Jones-Lee and the value of statistical life, health and safety¹

Graham Loomes^{1*} and Robert Sugden² 

¹Warwick Business School, University of Warwick, Coventry, UK and ²School of Economics, University of East Anglia, Norwich, UK

*Corresponding author: Graham Loomes, email: graham.loomes@wbs.ac.uk

(Received 22 September 2022; accepted 22 September 2022; first published online 1 December 2022)

Abstract

This is an appreciation of the life and work of Michael Jones-Lee. It describes his pioneering role in establishing and developing the theory and practice of the elicitation of monetary values for changes in risks to life, health and safety, using stated preference methods.

Key words: Value of statistical life; value of safety; stated preference methods

Michael Jones-Lee, who died in February 2021 at the age of 76, was a pioneer in the field of eliciting individuals' valuations of changes in risks of death and injury to provide guidance for public policy.

The 'value of life' or, more accurately, the value of statistical life (VoSL), also known as the value of a prevented fatality (VPF), was at the core of Mike's work from the late 1960s when, as a PhD student, he proposed a new methodology for inferring the value of life from survey data. In the early 1980s, working with psychologists on a project for the UK Department of Transport, he put this proposal into practice in the first large-scale survey-based investigation of individuals' valuations of reductions in risks of death. From then on, he was involved in almost every official study of this kind in the UK, as well as in a number of other countries, continually engaged in developing and refining the methodology and updating the values of life and safety used by government departments.

Stated preference studies were one of the earliest sources of evidence about how far individuals' preferences were consistent with the axioms of rational choice theory. From a very early stage in his work, Mike recognised that stated preferences sometimes showed patterns that contravened those axioms. Anticipating what has now become a major concern of behavioural economics, Mike looked for ways of improving survey designs to make them better able to elicit individuals' elusive 'true' preferences.

¹We thank Hazel Jones-Lee and Rupert Jones-Lee for information about Mike's early career and for insights about how he thought about his work.

This article is an appreciation of Mike's contributions to theory, evidence and policy. For both the authors, Mike was a long-standing friend, colleague and fellow researcher. Bob was taught by Mike as an undergraduate student in the late 1960s and was a colleague of his at the University of York in the early 1970s. In the late 1970s and for much of the 1980s, Mike, Graham and Bob were together in the economics department of Newcastle University. Graham worked with Mike on many stated preference studies of individuals' valuations of life and safety during several decades thereafter. All three of us have struggled with the problem of how to make public policy responsive to individuals' preferences when those preferences, as revealed in survey responses and in actual decisions, lack the consistency properties that economics has traditionally assumed.

From mechanical engineering to the value of life (via business finance)

Mike, who was born in 1944, began academic life as a student of engineering. Sponsored by Rolls-Royce, he spent a university apprenticeship year at their Derby works before beginning his degree course in Mechanical Engineering at the University of Sheffield in 1962, working at Rolls-Royce during the long vacations. Not surprisingly for a young man who had used his mother's dining room as a workshop for dismantling a motorbike (and who, in later life, loved and tinkered with his Austin-Healey Frogeye Sprite sports car and assembled a Caterham 7 to accompany it), he enjoyed working alongside the highly skilled craftsmen on the factory floor; but he was less at ease with the organisational culture that he was expected to join after graduating.

In his second year at Sheffield, Mike switched to a B.Eng. registration in Mechanical Engineering with Business Studies. One of his business finance tutors was A.J. (Tony) Merrett who, with Allen Sykes, had just published what was to become the principal British textbook on capital budgeting. Recognising Mike's exceptional ability, Merrett helped him to get a trainee position at a prestigious management consultancy in the US, with which Merrett had connections. After graduating with a First in 1965, Mike began work at the New York office of Joel Dean Associates.

Joel Dean was one of the great and the good of US business finance and managerial economics, the author of the classic texts *Capital Budgeting* (Dean, 1951a) and *Managerial Economics* (Dean, 1951b). It seems that Mike made an immediately favourable impression on him. Since a doctorate was deemed essential for progression as a consultant in the US, Dean proposed that Mike join the doctoral programme at Harvard Business School. At the age of 22, Mike was on the inside track for a high-powered career in business finance.

Around this time, however, Mike's father was developing serious health problems. Mike returned to Britain in 1966 to support his mother, and looked for a paid position that would allow him to study for a doctorate. With several offers on the table, he chose the one from the University of York, one of the new 'plate glass' universities of the 1960s. Its first professors of economics, Alan Peacock and Jack Wiseman, had high ambitions for their department, among which were the creation of an American-style graduate school with an international reputation. For this, they needed to recruit some first-class research students. Mike fitted the bill and was

offered a teaching fellowship alongside entry to the graduate programme. After completing the first year of the programme in 1967, he was registered for a D.Phil. and, still only 23, was appointed to a lectureship.

Mike's D.Phil. supervisors were Ron Cooper and Jack Wiseman, whose respective fields were statistics and applied economics. It seems that they left him a free rein to pursue his interests in pure economic theory, particularly the still relatively new topic of rational choice under uncertainty. This was a time when abstract mathematical theory was by far the most prestigious form of economics, and Mike's theoretical abilities were viewed with some awe. Initially, the topics he theorised about were taken from business finance and managerial economics. Significantly, these included portfolio selection. 'Some time during 1968' (according to the preface to his book *The Value of Life*), he attended a graduate economic theory seminar at which Alan Williams, then a Reader in the department, gave a paper about cost-benefit analysis (CBA). This, we are told, sparked off Mike's interest in the value of life and safety. From then, Mike must have worked very quickly. In January 1969, his first paper, 'Valuation of reduction in probability of death by road accident', was published in *Journal of Transport Economics and Policy* (Jones-Lee, 1969).

Mike completed his D.Phil. thesis in 1971 (Jones-Lee, 1971). The thesis has three sections, each containing three substantive chapters that would be more than sufficient for a present-day Ph.D. Section C, 'The private valuation of reduction in the probability of extreme outcomes', sets out an entire and radically new programme for evaluating changes in risks of death and injury. The theoretical core of this section, expressed in a condensed and elegant form, was subsequently published in *Journal of Political Economy* (Jones-Lee, 1974). Mike's later book, *The Value of Life: An Economic Analysis*, presents the same ideas as were in the thesis, but with more background material and a more substantial set of 'experimental results' (Jones-Lee, 1976).

The problem of valuing life

At the time Mike was working on his thesis, CBA was a thriving topic in economics. One of its most common areas of application, public decisions about road-building, called for monetary valuations of changes in the frequency of death and injury in road accidents. Thinking about how this should be done exposed unresolved questions about the meaning of 'cost' and 'benefit'.

There was a general understanding that CBA was a form of decision analysis, the public-sector analogue of capital budgeting. Beyond this, there was no consensus about the criteria it should apply. At least four different understandings of CBA were in an uneasy state of coexistence.

On one view, CBA was aligned with *national accounting*. The CBA of a project could be understood as measuring that project's net effect on national economic activity. On that view, a major element of the cost of a person's premature death is the resulting loss of *gross output* – what that person would have produced, had they lived a normal expected lifetime. This was the official position of the UK Department of the Environment (the government department then responsible for

road-building) in the early 1970s (Dawson, 1971).² This measure was conceptually coherent but ethically unpalatable. (For example, the premature death of a retired person resulted in no social cost.)

On another view, which Mike labelled *Olympian* (1976: 15), CBA was aligned with utilitarianism (or, more generally, with moral consequentialism). The underlying idea was that the role of a public decision-maker was to maximise an all-things-considered concept of the social good. Economists might legitimately propose methods of formalising social welfare rankings, but ultimately the definition of social welfare was a matter of ethical judgement, entrusted to the decision maker. This was broadly the position taken by two influential CBA manuals of the time (Dasgupta *et al.*, 1972; Little and Mirrlees, 1974). If this approach were to be applied to projects with life-and-death effects, the apparent implication was that CBA required all-things-considered ethical judgements about, quite literally, the value of a person's life. The issues involved in making such judgements did not fit easily into the economics-based framework of practical CBA.

A third view, favoured at the time by Williams (1972), aligned CBA with *output budgeting*.³ The fundamental idea was that rational decision-making was a matter of maximising an objective function subject to constraints. One of the primary roles of a public-sector decision analyst was to guide the decision maker towards specifying the intended outputs of the agency they directed and defining metrics for measuring those outputs. Maximising the agency's objective subject to a given budget constraint would then be a well-defined problem. Necessarily, the solution to this problem would imply shadow prices for units of output. When appraising individual projects, those shadow prices could be treated as measures of value. Thus, in the appraisal of road projects, rational decision-making might be held to require the use of some monetary VoSL, common to all appraisals. This approach was coherent and pragmatic, but it was a major retreat from what had long been seen as a primary task of economics – developing a *theory of value*.

On the fourth view, CBA was aligned with *welfare economics*, and in particular with what was still sometimes called *new* (i.e. post-utilitarian) welfare economics. This was the position taken by Mishan (1971a) in an influential textbook of CBA. The core idea was that CBA was an application of the *potential Pareto improvement* criterion (or hypothetical compensation test). A project passes this test if, in conjunction with some hypothetical scheme of lump-sum transfers of income, it would make some people better off (in terms of their own preferences) and make no one worse off. It was well known that, under standard assumptions about individuals' preferences, there can be states of affairs *A* and *B* such that a proposed move from *A* to *B* clearly fails to satisfy the potential Pareto improvement test, but the same is true of a proposed move from *B* to *A* (the 'Scitovsky paradox'). If a project has relatively small effects on the income of each affected individual, this problem can be safely assumed

²Previously, and even less defensibly, UK governments had used a measure of *net* output, i.e. the foregone output of the dead person minus the value of what they would have consumed, had they lived (Reynolds, 1956; Dawson, 1967).

³In a later textbook, Sugden and Williams (1976) presented this understanding of CBA in parallel with the approach we describe in the next paragraph, maintaining a stance of neutrality.

away in an appraisal, but that is clearly not the case if someone has to be compensated for their own death or serious injury. If costs and benefits are to be assessed in terms of a project's *ex post* effects, the problem seems inescapable.

Mike's proposal

Mike's proposal cut through these difficulties and ambiguities with panache. He made three bold moves, which we consider in turn.

The foundations of CBA

Aligning himself with what at the time was a minority position among cost-benefit analysts, Mike endorsed the potential Pareto improvement test as the criterion to be used in CBA. In doing this, he offered two cogent supporting arguments.

The first was a matter of consistency. In practice, almost all CBAs make implicit appeal to the principle of consumer sovereignty when they use measures of consumers' surplus to value consumption of goods that are bought and sold in markets. Consistency requires the same principle to be used when evaluating changes in the provision of public goods:

The existence, or otherwise, of a market simply determines the ease with which the economist may obtain information concerning the nature of underlying preferences which are generally assumed to exist prior to, and independently of, media for their expression. (Jones-Lee, 1971: 279)

The second argument expressed Mike's conviction that the methods of CBA should be grounded in rigorous theory and be independent of the judgements of particular decision makers. If CBA uses the potential Pareto improvement criterion, the analyst does not need to 'involve himself intimately in the views and ideology of the social decision-maker'. This gives the results of CBA greater generality and makes it more likely that a 'coherent and consistent body of analytical technique' will emerge (Jones-Lee, 1976: 15). If CBA is to have this kind of scientific independence, there can be no claim that public decisions ought to be fully determined by CBA. The results of such an analysis must be understood as 'only *part* of the information relevant to any particular allocative decision' – but the part to which economics can most usefully contribute (Jones-Lee, 1976: 16, emphasis in original).

Valuing statistical life

Mike's second move revealed his brilliance as a theorist. Introducing Section C of his thesis as a continuation of the preceding chapters on business finance, he declared that the fact that safety is often a public good

... should not be allowed to obscure the essential similarities (from the consumer's point of view) between a change in the probability of occurrence of an extreme outcome and a change in the mean or variance of returns from an

asset portfolio. Both represent, essentially, changes in the parameters of probabilities of outcomes between which the individual has a (more or less) well-defined set of preferences. (Jones-Lee, 1971: 278)

The crucial step was to recognise that the ‘lives’ that CBA needed to value were *statistical* lives. For the projects typically appraised by CBA, the relevant effects are relatively small changes in any individual’s risk of death or injury, but often impacting on many individuals. Given the preference-based principles of the potential Pareto improvement criterion, what is required is a measure of each individual’s compensating variation for changes in risk. Mike saw the development of such a measure as a problem in pure theory, analogous with the formal development of demand theory. The theory of demand derives qualitative implications about the properties of individuals’ demand functions from very general assumptions about individuals’ preferences over consumption bundles. Mike looked for qualitative implications about compensating variations for risk that could be derived from expected utility theory (EUT).

The simplest version of Mike’s analysis is a one-period model in which an individual faces some probability of death (Jones-Lee, 1976: 66–67, 92–99). Possible consequences are represented by ordered pairs (w, h) , where w is the individual’s level of wealth at the start of the period and h is a binary variable which can take the value *life* (the individual survives the period) or *death* (does not survive it). EUT implies that if the individual’s preferences over uncertain prospects satisfy certain axioms, each such consequence has a utility value, unique up to positive linear transformations. Mike represents these utilities using two functions, $L(w)$ and $D(w)$, conditioned respectively on life and death. Crucially, these functions are representations of the individual’s *ex ante* preferences over prospects; they are not to be interpreted as *ex post* measures of the value of being alive or dead. Mike considers an individual with an initial wealth \bar{w} and an initial probability \bar{p} of dying during the period. He then derives the formula for the individual’s *compensating variation* $v(p)$ for a change in the probability of death from \bar{p} to p . He assumes that individuals are risk-averse, prefer more wealth to less (strictly so when conditional on life, weakly so when conditional on death) and, at given levels of wealth, prefer life to death. Following a standard practice in neoclassical economics, he assumes that $L(\cdot)$ and $D(\cdot)$ are twice differentiable. Using the mathematics of differential calculus, he shows that $v(p)$ is a smooth and concave decreasing function of p with $v(\bar{p}) = 0$. In the range $p > \bar{p}$, $-v(p)$ measures the monetary compensation sufficient to induce the individual to accept additional risk; in the range $p < \bar{p}$, $v(p)$ measures the individual’s willingness to pay for a risk reduction.

The concavity of the $v(p)$ function implies that, for any given finite increment of probability Δp , the individual’s willingness-to-pay (WTP) valuation of a Δp reduction in risk is lower (in absolute value) than their willingness-to-accept (WTA) valuation of a Δp increase in risk. But because the function is smooth, its first derivative at $p = \bar{p}$ is well defined. The negative of this derivative is the *marginal value of a decrease in risk*. This is the value that is appropriate for the CBA of projects that have small effects on the safety of individuals. That the same marginal value applies to small increases and small decreases in risk implies that, when changes in risk are

small, the Scitovsky paradox is not a problem for CBA. Suppose a project reduces the probability of death in the next period by $1/n$ for each of n individuals, so that the project is expected to 'save' one 'statistical' life. Suppose also that n is large and that the i th individual's marginal value of a $1/n$ decrease in their own risk of death is v_i . The sum of the marginal values v_i across the n individuals can be interpreted as their collective *value of a statistical life*.

In parallel with this one-period model, Mike analyses a continuous-time model in which risk is measured in terms of the individual's life expectancy μ . The individual's compensating variation for an increase in life expectancy from $\bar{\mu}$ to μ can be represented by a function $v(\mu)$. This is a smooth and convex increasing function with $v(\bar{\mu}) = 0$. The first derivative of this function at $\mu = \bar{\mu}$ is the *marginal value of an increase in life expectancy*.

The basic idea of valuing statistical life in terms of individuals' preferences over uncertain prospects pre-dates Mike's work. The first statement of this idea was probably by Drèze (1962), who used it in an exercise in social choice theory.⁴ The idea of using the VoSL in project appraisal was proposed by Schelling (1968) in a paper entitled 'The life you save may be your own'. The same idea was later advocated by Mishan (1971b), who cited Schelling but was apparently unaware of Mike's 1969 paper. Nevertheless, Mike can be credited with the first formal theory of the VoSL.

Elicitation of preferences

Mike's third move was perhaps the boldest of all. If CBA is to use the potential Pareto improvement criterion, there must be some operational method by which an analyst can discover individuals' compensating variations for the effects that need to be valued. When those effects are changes in the market prices of private goods, the theory of consumers' surplus shows how compensating variations can be inferred from observable demand functions. But what if the effects are changes in the provision of public goods, or in the consumption of goods that are not traded on markets? The prevailing view among cost-benefit analysts of the time was that inferences about individuals' preferences should be based on evidence from actual choices. For example, estimates of individuals' valuations of travel time savings might be inferred from data about choices between alternative travel modes; estimates of individuals' valuations of reductions in air pollution or aircraft noise might be inferred from econometric analyses of property values. Although the first attempts to infer valuations of risk reduction from choice data – usually data about wage differentials – were published after 1971 (Smith, 1973; Thaler and Rosen, 1973; Melinek, 1974), Mike recognised this as a possibility in his thesis (1971: 323–325). He discussed some of this work in *The Value of Life* (1976: 38–40, 43).⁵

⁴Drèze was a formal theorist in the style that Mike most admired, and Mike always recognised Drèze's 1962 paper as seminal. A high spot in Mike's early career was presenting his work on the value of life at an International Economics Association Theory Workshop organised by Drèze and James Mirrlees in 1971.

⁵In an unpublished 1977 conference paper, Mike reported his own attempt to use this method to infer a value of statistical life from data about the frequency with which motorists replace car tyres (see Jones-Lee, 1989: 68–69).

Dissenting from the received wisdom, Mike proposed what he described as an ‘experimental procedure’ of ‘direct enquiry’ – that is, asking representative samples of individuals to report their own preferences. Responding to the objection that introspection is not ‘universally testable’ and thereby not a legitimate source of data for science, he described that position as ‘extreme’, ‘arbitrary’ and ‘nihilistic’, especially in cases where no other satisfactory data are available (Jones-Lee, 1976: 130). The real problem for the experimental method in the social sciences, he maintained, was that of isolating the effects of the variables under investigation by controlling other variables. In some cases, direct enquiry is more controlled than methods based on actual choices, whether those choices are made in real-world settings or in incentivised laboratory tasks:

‘Objective’ market data are usually every bit as much the result of *uncontrolled experiment* as the answers to a consumers’ survey asking hypothetical questions about ‘how much, or what, one would buy if ...?’ Neither source of data is ideal as a basis for inferences concerning, for example, the demand relationship between quantity and price. (Jones-Lee, 1971: 325–326, emphasis in original)

The implication is that the fundamental concept in decision theory is not *choice* (as in revealed preference theory); it is *preference* itself. Choice behaviour is one source of information about a person’s preferences, but it does not have automatic privilege over other sources. Mike may have had in mind the role of probability judgements in decision-making. In a survey question, it is possible to control respondents’ beliefs by including relative frequency information in the description of a hypothetical scenario. In contrast, when real market choices are used to elicit preferences, unobserved variation in individuals’ beliefs is a potential source of loss of control.⁶

Mike acknowledged two potential problems with the method of direct enquiry. First, if individuals are asked to respond to unfamiliar hypothetical scenarios, they may be unable to imagine how they would behave if the scenario were real. Second, if individuals know how their responses will be used by the cost–benefit analyst, they may have incentives to misrepresent their preferences. Mike argued that these problems could be mitigated by careful survey design – by using scenarios that respondents find easy to imagine and by not drawing attention to the ultimate object of the survey (Jones-Lee, 1971: 323–325, 1976: 128–130).

Here too, Mike’s position was aligned with Schelling’s 1968 paper. Schelling had argued that direct enquiry was a valid method of eliciting individuals’ valuations of safety, and had sketched a ‘scaling’ or ‘chaining’ method by which valuations of very small changes in the probability of death could be elicited by using two more easily understandable decision problems in combination – a method later used by Mike. Mishan also recognised that survey methods might be useful in eliciting

⁶In presenting expected utility theory in *The Value of Life*, Mike treats preference and subjective probability as primitive concepts. Formally, the theory is about the properties of preferences over ‘subjective probability distributions’. In presenting his survey method, he makes the explicit assumption that respondents’ subjective probabilities match the relative frequency information they have been given (Jones-Lee, 1976: 56–66, 133–134).

valuations of safety, but contented himself with the remark that, while cost–benefit analysts can observe the quantities that individuals buy at market prices,

... one cannot generally observe their subjective valuations. In the circumstances, economists seriously concerned with coming to grips with the magnitudes may have to brave the disdain of their colleagues and consider the possibility that data yielded by surveys based on the questionnaire method are better than none. (Mishan, 1971b: 704–705)

Mike had the bravery, not merely to advocate the questionnaire method, but to use it. In the final chapter of his thesis and in his 1969 paper, he reported an ‘illustrative exercise’ in which he had posed survey questions to one individual and used the responses to elicit his $v(p)$ function (Jones-Lee, 1971: 329–351).

In *The Value of Life*, Mike reported a much more substantial but still ‘illustrative’ survey, designed to elicit respondents’ $v(p)$ and $v(\mu)$ functions. He sent questionnaires to a convenience sample of around 90 people, of whom 31 (almost all academics or research workers) responded. The $v(p)$ questions were about a hypothetical scenario in which the respondent had to choose between flying by two alternative airlines, each with a specified ‘recent safety record’ of fatal crashes per 500,000 flights. This safety record differed across questions. Respondents stated reservation prices for choosing ‘airline B’, given the price charged by ‘airline A’. The $v(\mu)$ questions were about a hypothetical scenario in which the respondent had to choose between two alternative job locations which differed only in terms of house prices and ‘the impact of environmental pollution on life expectancy’. Area B had a specified increment (positive or negative) of life expectancy relative to the ‘normal’ level of area A. Respondents stated reservation ‘premia’ or ‘discounts’ on house prices in B relative to A.

Mike reports the full set of responses by each of the 31 respondents. On the basis of various simplifying assumptions (and a certain amount of hand-waving, understandable in the context of an illustrative exercise), he concludes that the $v(p)$ data imply a mean value of statistical life of £3.1 million and that the $v(\mu)$ data imply a mean value of life expectancy of £900 per year (Jones-Lee, 1976: 142).⁷ He sketches $v(p)$ and $v(\mu)$ graphs for six of the respondents and concludes that these tend to confirm his theoretical analysis of the general properties of those functions (Jones-Lee, 1976: 143–145). By this, he means that if smooth curves are fitted to the individual-level data, these curves are generally decreasing for $v(p)$, increasing for $v(\mu)$, and concave in both cases.

However, some features of these responses are not easy to reconcile with EUT. The smoothness of the $v(p)$ and $v(\mu)$ curves – the property that ensures that marginal WTP and marginal WTA are equal, neutralising the Scitovsky paradox – is assumed rather than inferred from the data. In fact, the data raise doubts about this assumption. For example, the median WTP valuation of an additional year of life expectancy (the smallest increment for which responses are elicited) is £200, while the median

⁷Approximately equivalent to £27 million and £8000, respectively, in 2021 prices.

WTA valuation of a loss of year of life expectancy is –£500.⁸ And there is an apparent discrepancy between the two estimates of the VoSL. Given that none of the respondents had a current age-specific life expectancy greater than 60 years, £900 per year of life expectancy converts to less than £54,000 per statistical life.⁹

In carrying out this survey, Mike had made a serious attempt to do what Schelling and Mishan had only speculated about. Clearly, however, he was feeling his way in unfamiliar scientific territory. Almost all his experience as an economist had been in tackling abstract theoretical problems. By 1976, he had published 17 journal articles, of which 16 were entirely theoretical. (The exception was the 1969 paper, which had included his one-respondent survey.) Of course, the survey method was new to him, as it would have been to almost any other economist of the time, but he had little experience of any kind of empirical research. It is noticeable that, in drawing inferences from the survey data, Mike used no econometric estimation methods or statistical tests. In publishing *The Value of Life*, he put himself in an unusual position. He was advocating a research method that most economists viewed with scepticism or even hostility and for which his previous career as a theorist provided little preparation.

Raising the empirical stakes

Having nailed his colours to that mast, Mike took steps to extend the scope and scale of his empirical investigations. He established contact with officials at the Department of Transport (DTp) and persuaded them that a well-designed survey of people's attitudes to transport risks had the potential to provide them with a great deal of relevant information about public preferences, including guidance about the money values that might be placed on changes in risks to life and limb.

After moving to Newcastle University in 1977 to take up a Chair in Economics, Mike began a collaboration with colleagues from the Psychology Department (Max Hammerton, a professor there, and Vicki Abbot, a graduate student) and the Health Care Research Unit (Ian Russell, a medical statistician, and Peter Philips, a data analyst).

DTp agreed to fund a two-phase programme, consisting of an initial feasibility study which, if sufficiently encouraging, would lead to a national representative sample survey. The principal features of the design and results of the feasibility study are reported in Hammerton *et al.* (1982a) with some supplementary analysis reported in Hammerton *et al.* (1982b).

The coming together of an economic theorist armed with a model of an idealised rational individual and a psychologist well-versed in experimental techniques for studying actual human responses resulted in a substantial advance in stated preference methodology in the field of risk and safety.

⁸It is not possible to make a similar comparison for the airline safety data because, in the survey design, the increments of increases in risk do not match the increments of reductions.

⁹In Jones-Lee (1989: 79), Mike suggests that this discrepancy might be explained by the huge difference in scale between a year of life expectancy (the smallest unit used in the questionnaire study) and the risks of death involved in single airline flights. That is psychologically plausible, but the discrepancy is still contrary to EUT.

The design of the feasibility study was both more ambitious and more sophisticated than any stated preference VoSL studies up to that time. Although the 120 participants in that study constituted a convenience sample (all drawn from the vicinity of Newcastle), they were more diverse than in most of the previous small studies, with recruitment aimed at producing gender, age and income distributions that were not too dissimilar from the population at large. Participants were randomised to some degree between different versions of questions and elements were built in to check for the consistency of answers with basic assumptions of the economic model. There were also checks for the extent to which answers were consistent across different question formats; and the stability of responses was examined by re-interviewing 30 of the 120 participants about a month later.

All 120 interviews, as well as the 30 re-interviews, were conducted by Vicki Abbott (later Abson) who was able to provide feedback about how well or poorly different formats or different types of questions were comprehended. Indeed, some of the questions were specifically aimed at exploring people's understandings and perceptions of probabilities and outcomes, as well as their views about the acceptability of putting monetary values on personal safety.

Having satisfied themselves – and the DTp funders – that members of the general public were willing and able to engage with the kinds of questions involved, the researchers moved on to the second phase, commissioning an experienced survey organisation – National Opinion Polls (NOP) – to undertake the study with a nationally representative sample. The key features of this 'main' study are summarised in Jones-Lee *et al.* (1985), with a fuller and more detailed account and discussion appearing as a co-authored Chapter 4 in Jones-Lee (1989).¹⁰

Many of the issues and challenges that Mike engaged with in much of his later work make an early appearance in the DTp studies. How could one formulate questions that would present scenarios involving the key features – small changes in small probabilities of death or injury over specified periods, to be traded off against sums of money – while making them reasonably simple, unambiguous and tolerably realistic? How could one get at the relative values of different levels of severity of non-fatal injuries compared with fatalities? Were different causes of death valued differently by the population? Were reductions in risks treated very differently from increases in risks? How did people handle information about the small probabilities involved: did different ways of expressing/representing those probabilities affect the way they were understood and did people's responses to probability manipulations accord with the axioms incorporated into standard economic theory? Did different instruments for eliciting money value responses systematically affect the size and sensitivity of the answers given? In short, how confident could one be about the extent to which (most) people's preferences were structured more or less in accord with the model of the rational agent?

In order to provide 'quality assurance' checks, the researchers incorporated several basic consistency tests, asked risk-wealth trade-off questions in a number of different ways and revisited a subsample of 210 respondents to examine the stability of responses. Taken together, the two phases of the DTp study set high standards at

¹⁰The authorship of this chapter is given as P.R. Philips, I.T. Russell and M.W. Jones-Lee.

this early stage in the evolution of stated preference attempts to produce values of statistical life.

Table 4.30 in the 1989 book chapter provides a useful ‘summary of the various results enhancing and detracting from the credibility of estimated marginal rates of substitution’.

On the credit side, the authors suggested that there was a respectable, if imperfect, degree of consistency with basic requirements of rationality and that there were broad similarities in the distributions of responses to the two questions they considered most pertinent (Q18 and Q20),¹¹ that central tendency measures varied with risk changes in the expected direction and that coefficients from regression estimates could be reconciled with the underlying theory.

On the debit side, in response to a question (Q17) testing for basic understanding of probabilities, 48% gave the correct answer but 47% gave the wrong answer (the other 5% didn’t know). That question was as follows:

‘Imagine that you have to face two different risks of being killed:

In one, your risk of death is 2 in 100,000

In the other, your risk of death is 20 in 100,000

You cannot avoid either of these risks but you can choose to have one of them reduced. Which would you prefer:

The risk of 2 in 100,000 reduced to 1 in 100,000

The risk of 20 in 100,000 reduced to 15 in 100,000’

If understood and processed according to standard assumptions, this is a choice between reducing the risk of death either by 1 in 100,000 or else by 5 in 100,000, so that choosing the first option rather than the second seems like a clear (and substantial) error. This highlights one of the challenges that continues to face researchers working in contexts involving small changes in already small baseline probabilities: namely, many people are not familiar with explicit formulations of such data and their ways of perceiving and processing these numbers may not be as professional statisticians and decision theorists would suppose. Arguably, however, some of the subsequent questions upon which the researchers based their main conclusions did not involve more than one baseline at a time, and for those questions the NOP interviewers were equipped with visual aids depicting the relevant probabilities in terms of ‘blacking out’ squares on pieces of graph paper containing 100,000 small squares.

Another issue discussed in the 1989 chapter concerns the particular method for eliciting a WTP amount. As the feasibility study had indicated, and as the main study was to confirm, many individuals’ responses were liable to diverge considerably – but seemingly, unsystematically – between their initial interview and the re-interview about a month later. This suggested – and many other studies have reinforced this suggestion – that there was a considerable degree of uncertainty in people’s minds about their valuations.

¹¹Q20 is discussed in more detail below. Q18 used a scenario involving different coach services with different safety records and aimed at eliciting WTP and WTA values for different levels of risk relative to a reference scenario.

The researchers drew some comfort from the fact that the test–retest variability appeared to be unsystematic: with a sufficiently large representative sample, it might be hoped that the ‘noise’ would cancel out and that the central tendency measures might be relatively reliable. At the same time, they were aware that respondents’ uncertainties might make them susceptible to influence from the particular system of prompting that was used, but felt that the best that could be done in this study was for the same system of prompts to be used consistently across all questions.

To illustrate how Mike and his colleagues explored the extent to which patterns of response were consistent with the implications of the theory, consider Question 20. This might be considered to be a ‘prototype’ question, in the sense that it used a scenario that goes directly to road safety and that was adapted and refined in later studies in which Mike was prominent. Briefly, it asked how much the respondent would pay for an extra safety feature to be fitted to a car they were thinking of buying. The baseline risk, without the extra safety feature, was given as 10 in 100,000 (per year). Q20(a) asked for WTP for a device that would ‘halve the risk of the *car driver* being killed, down to 5 in 100,000’ (italics in the original, here and below). Q20(b) asked about an alternative device that ‘would reduce the risk of the *driver* being killed by 20% to 8 in 100,000’. And Q20(c) asked for WTP if the device ‘would halve the risk of the *driver* and *any passengers* in the car being killed’ (with the interviewer prompting the respondent to think in terms of an average of one passenger).

On the basis of the theoretical model Mike had propounded, one might have expected the answer to part (a) to be at least double the answer to (b) – even more than double under some assumptions. One might also have expected a substantially higher amount in part (c) compared with part (a), especially where the additional passenger(s) were likely to be members of the respondent’s family.

However, actual responses were not always so well-behaved, in several respects. First, some people may have found the numbers challenging. For whatever reason, 11% of the sample said they were willing to pay *more* for the smaller risk reduction. Second, even though one might suppose a reduction in the risk of death to be worth at least something, a proportion of respondents gave zero responses: for parts (a), (b) and (c) of Q20, those proportions were 13%, 18% and 10%, respectively.

A third subset appeared to acknowledge that reducing the risk of death was worth paying something for, but did not differentiate according to the magnitude of the benefit. So, in addition to the 13% who gave the same zero response to parts (a) and (b), there were another 34% who gave the same non-zero answers to both parts. Many individuals’ answers to part (c), when compared with their part (a) answers, also displayed a lack of sensitivity to the addition of passenger(s), with the median difference (interpreted by the theory to be the median value placed on halving passenger risks of death) being zero.¹²

When many individuals’ responses exhibit less sensitivity to key variables than the theory supposes, the consequence is that the estimated VoSL is liable to differ systematically, depending on the size of the risk change(s) selected by the creators of the

¹²In fact, 14% of the sample gave a *lower* WTP for part (c) than for part (a) which, if not interpreted as error/misunderstanding, would imply a negative value of passenger safety, while another 47% gave the same answers to both parts.

scenarios, with smaller risk reductions associated with larger estimates of VoSL. So, the VoSL based on the trimmed mean of responses to part (b) – £2.21 m in 1982 prices – was more than 80% higher than the corresponding figure – £1.21 m – inferred from the answers to part (a).¹³

However, this degree of variability seemed modest when compared with the extent to which values varied when derived from rather different question formats. Thus, Question 14, formulated in terms of reducing annual deaths by 100, when applied to the subset focusing on transport fatalities, generated a trimmed mean VoSL (for self plus others) of £7.35 m, while Question 22(b), couched in terms of paying higher local taxes, produced a trimmed mean based figure of £280,000: that is, the Q14 format resulted (under the various assumptions made) in a VoSL more than 25 times bigger than the value inferred from Q22(b).

Mike and his collaborators did not shy away from these results and the challenges they posed. Indeed, it was their willingness to incorporate numerous tests for consistency and coherence that brought them clearly out into the open. The question was: how to react?

Defending the VoSL methodology

The fullest statement of Mike's approach to the valuation of life and safety is his 1989 book, *The Economics of Safety and Physical Risk* (Jones-Lee, 1989), which revises, expands and updates *The Value of Life*. It includes a long chapter about the DTP project (discussed in the previous section). Because we have wanted to emphasise the originality of Mike's contributions, we have focused on his thesis and on his 1976 book. However, the later book contains some significant reflections on issues that came to prominence after 1976, and on Mike's experience of working with psychologists on stated preference research.

One post-1976 development was a decade-long debate between Mike and the philosopher John Broome, who criticised the VoSL approach from what Mike characterised as an Olympian perspective (Broome, 1978, 1985; Jones-Lee, 1979, 1987). Broome argued that, if social decisions about risks of death were based on individuals' compensating variations for risk reductions, those decisions could violate conditions of collective rationality. For example, the sum of compensating variations for a 1/100 risk of death to each of 100 individuals would be greater than the corresponding sum for a 1/1000 risk to each of 1000 individuals. (This is an implication of the concavity of the $v(p)$ function.) Broome argued that the two risk changes should be valued equally, because their expected outcomes (i.e. one death) were exactly the same. After pointing out that this was not a serious problem in practice because policy-relevant risks to individuals are usually very small, Mike responded to the issue of principle, saying:

[I]t is clear that under certain circumstances the dictates of coherence and consistency in government decision-making will inevitably conflict with

¹³A similar effect may explain the apparent disparity (discussed above) between responses to the airline safety and life expectancy questions in the 'illustrative survey' reported in *The Value of Life*. See footnote 9.

considerations of democracy (widely construed to include a requirement that government decisions should take account of individual wishes and attitudes to risk). (Jones-Lee, 1989: 20)

For Mike, 'democracy' (in the sense of respect for individuals' preferences) was a more fundamental principle of collective decision-making than coherence.

Broome, like many philosophical critics of the WTP approach, distinguished between preferences and interests. He argued that CBA should be based on individuals' interests, and that preferences are not reliable indicators of interests. Mike's response upheld the principle of democracy, but with a significant proviso:

[G]iven adequate information concerning the likely consequences of decisions affecting safety, the majority of people would, I suspect, prefer that such decisions should reflect their own 'true' preferences and attitudes to risk (i.e. the preferences and attitudes that they would arrive at on careful reflection rather than in the heat of a rapidly made and possibly fickle decision). (Jones-Lee, 1989: 19)

Here, Mike is invoking a concept of *true* preference, defined as the preference that an individual would report in the light of adequate information and after careful reflection. He maintained that the survey method has the potential to elicit such preferences if correct and explicit information is built into the questions and if the questions are 'readily intelligible, believable, unambiguous and, so far as possible, involve choices of a type that are familiar to the respondent' (1989: 74). These are properties of valuation questions that Mike tried to satisfy in all his survey designs. As a gold standard, Mike proposed assembling representative panels of individuals in settings similar to focus groups and asking valuation questions only after a period of instruction and discussion, thus facilitating well-informed and careful reflection (1989: 289).

It was a fundamental assumption of the VoSL approach that individuals have well-articulated preferences that are broadly consistent with EUT. By 1989, however, economists were becoming aware of a growing body of experimental evidence of regularities in human decision making that contravened EUT – regularities such as the common consequence and common ratio effects, preference reversal, scale and scope insensitivity and WTP-WTA disparities. As noted in the previous section, the DTp project had found some apparently systematic inconsistencies in individuals' responses. Having worked with psychologists on that project, Mike was more conscious than most economists of the time of the potential significance of psychological factors in explaining economic behaviour. In the 1989 book, he offered a justification of the VoSL approach that acknowledged the findings of cognitive psychology and experimental economics while maintaining the fundamental validity and usefulness of EUT.

If the VoSL approach is to work, there has to be some bridge between the specific, highly stylised and hypothetical decision problems presented to survey respondents and the public decision problems that are to be subject to CBA: survey responses must be organised in relation to some theoretical framework. Since the aim is to elicit preferences with respect to *risks* of death and injury, that framework must be one of

choice under uncertainty. Mike argued that, all things considered, EUT remained the best all-round theory of choice under uncertainty: '[I]t remains the author's firm conviction that, as a simple summary description of the way most people tend, on the whole, to make choices under uncertainty, the Von Neumann-Morgenstern or Savage axioms retain a powerful appeal' (Jones-Lee, 1989: 105). And: '[N]o alternative theory of individual choice under uncertainty so far proposed is clearly superior on all, or even a majority of, the criteria by which such theories might be judged' (Jones-Lee, 1989: 106). In supporting this claim, Mike invoked the criteria of generality, parsimony, simplicity, predictive power and the intuitive appeal of a theory's axioms.

Mike recognised that psychological theories could explain some aspects of individual choice that EUT neglected, and characterised economics and psychology as complementary disciplines:

[T]he contributions from psychology ... are most fruitfully viewed as complements to, rather than as substitutes for, the approach adopted by economists. ... It is rather as if economists, concerned with a rather narrow range of questions related to the problem of resource allocation, focus a spotlight on a limited area, while psychologists play a less intense beam, more widely cast. (Jones-Lee, 1989: 237–238)

Mike accepted that, in some cases, the effects picked out by the broader beam of psychology were properties of 'true' preferences. For example, psychological research had shown that individuals' attitudes to objectively equivalent risks differed according to the source of the risk: people were more tolerant of familiar risks than of those that were perceived as 'dread' or 'unknown' (Slovic *et al.*, 1981). According to Mike, such effects could and should be taken into account in VoSL measures:

[I]f people are, for whatever reasons, more averse to risk from nuclear power generation than to an 'objectively' equivalent risk from, say, a natural disaster, then this comparative aversion should be reflected in decisions concerning the allocation of resources to reactor safety and disaster prevention and this is precisely what the economic approach seeks to do. (Jones-Lee, 1989: 224)

However, he was more sceptical of claims that experimental evidence had shown systematic *general* violations of fundamental EUT axioms. Using the 'common ratio' version of the Allais Paradox as an illustration of such alleged effects, he concluded:

[T]he kind of choices in which people have been found to exhibit systematic violations of the axioms tend to be rather special and indeed, one might be forgiven for suggesting, somewhat contrived. ... It could therefore be argued that paradoxes of this type pose no more of a threat to expected utility theory *as a summary description of general tendencies in choice under uncertainty* than do the standard optical illusions to the proposition that most people have, on the whole, quite acute powers of visual perception. (Jones-Lee, 1989: 105–106, emphasis in original)

Mike restated his 'firm conviction' that, by virtue of the concavity of the $v(p)$ function, EUT provided 'a perfectly adequate explanation for the difference between observed WTP and WTA in the context of physical risk' (Jones-Lee, 1989: 227).

In summary, the Mike of 1989 was conscious of the problems involved in reconciling the VoSL approach with the findings of behavioural psychology. His response was pragmatic. He held to the belief that the theoretical model, although idealised, was a reasonable approximation to the underlying structure of the preferences that people would express if adequately informed and after reflection, and that the task ahead was to develop finer instruments and techniques for extracting sufficiently robust core values. And by and large, that was the guiding principle that shaped much of Mike's research agenda for the rest of his life.

Moving beyond direct valuation methods

The next opportunity to try to develop those finer instruments and techniques came when the DTp, in conjunction with its Transport Research Laboratory (TRL), commissioned studies to assign values to reductions in the risks of *non-fatal injuries* which could be used alongside the VoSL in transport project appraisal. A detailed account of the main study is provided in Jones-Lee *et al.* (1993).¹⁴ Here, we focus on the key features and main conclusions drawn.

Given some of the concerns about the sensitivity and reliability of respondents' answers to the kind of valuation questions used in the earlier main VoSL study and in other piloting work, it was decided to use two distinct questionnaires. The *contingent valuation* (CV) questionnaire used broadly the same approach as in the VoSL study, but with risks of injury substituted for risks of death. That is, it directly elicited respondents' WTP for specified reductions in risks of different injuries. The *standard gamble* (SG) questionnaire did not refer to amounts of money, but instead elicited respondents' trade-offs between relatively minor injuries with certainty and risks of more serious ones. Because one of the possible outcomes was death, this approach was a way of eliciting respondents' valuations of reductions in injury risks *relative to their valuations of reductions in risks of death*. Since the DTp already had a VoSL, such relative valuations would be sufficient for its purposes.

To enable comparability between the two approaches, all interviewers were trained to administer both questionnaires and respondents were randomised between the two. The stratified sample was considered to be acceptably representative of the population and resulted in 414 completions for CV and 409 for SG.

The study as a whole revolved around six states of health,¹⁵ each depicted on separate showcards labelled with a letter of the alphabet. Four of these cards showed the (stylised) consequences of different severities of injuries arising from non-fatal road

¹⁴The main study was preceded by a feasibility study, parts of which appeared in journal articles (Dubourg *et al.*, 1994; Dolan *et al.*, 1995). A different study by researchers at the University of East Anglia, using a somewhat different approach, was also commissioned to run at the same time: see O'Reilly *et al.* (1994) for an overview.

¹⁵Actually, there were ten states, but four did not feature after the initial stages of the questionnaires. The full descriptions of all injury scenarios are given in Jones-Lee *et al.* (1993).

accidents, giving information about time in hospital, degree of pain/discomfort and duration of impact on subsequent activities. Briefly, they can be summarised as:

- W – less severe, full recovery 3–4 months;
- X – moderately severe, full recovery 1–3 years;
- S – somewhat more severe than X, with *some* long-term impairment;
- R – most severe of the four, with significant permanent disability.

The other two cards were: J – ‘Your normal state of health’ and K – ‘Immediate unconsciousness, followed shortly by death’.

Common to both questionnaires was an initial section asking respondents to rank all cards from least bad to worst and then to locate them on a visual analogue scale (VAS) ‘thermometer’ with the health state ranked best assigned a score of 100 and the state ranked worst assigned a score of 0. Both types of questionnaire then explained how risks were going to be expressed in subsequent questions and conducted some basic checks for comprehension. After that point, the two questionnaires diverged.

The CV questionnaire presented respondents with a series of six questions where the basic scenario was the same: a safety feature to be fitted to the respondent’s car which would be effective for a year in reducing the driver’s chances of suffering a specified injury by a specified amount. The six specifications, together with the codes assigned to them, were:

- K1:** risk of death (K) reduced by 4 in 100,000 from 8 in 100,000 down to 4 in 100,000;
- R1:** risk of injury R reduced by 4 in 100,000 from 16 in 100,000 down to 12 in 100,000;
- S2:** risk of injury S reduced by 12 in 100,000 from 24 in 100,000 down to 12 in 100,000;
- S1:** risk of injury S reduced by 4 in 100,000 from 24 in 100,000 down to 20 in 100,000;
- X2:** risk of injury X reduced by 12 in 100,000 from 30 in 100,000 down to 18 in 100,000;
- W2:** risk of injury W reduced by 12 in 100,000 from 16 in 100,000 down to 4 in 100,000;

For each of these risk reductions in turn (always in the order above), respondents were presented with a list of 22 amounts from 0 to £500 (plus ‘more than £500’) and were asked to put a tick against all amounts they were sure they *would* pay, a cross against all amounts they felt sure they *would not* pay, and an asterisk against ‘the amount which you think is your best estimate of what you would be willing to pay for that annual risk reduction’.

The issue that emerged from the CV part of the study was that many respondents’ answers to these questions appeared to show much less sensitivity to the differences between the risk reduction scenarios than might have been expected, even when one of the components – either the severity of injury or else the size of the risk reduction – was held constant.

For example, the first question asked for WTP to reduce annual risk of death by 4 in 100,000 and the next question asked for WTP to reduce the risk of injury R by the same amount, 4 in 100,000. Three hundred and forty-two respondents had said, in the earlier ranking and VAS exercises, that they regarded death as worse than R, so one might have expected them to give a lower WTP for the same size risk reduction for the less bad outcome. But in fact, 131 (38.3%) of them gave exactly the same non-zero amount, even though the R scenario came immediately after the death scenario and it would have been easy to adjust the response downwards. Meanwhile, another 43 (12.6%) who ranked injury R as less bad than death actually reported a strictly higher WTP to reduce the risk of R by the same amount.

Similar patterns applied to other pairings of injuries. Just as (if not more) discouragingly, there was substantial insensitivity to the risk reductions. Notice that the third and fourth questions involve the same injury, S, but different risk reductions: 12 in 100,000 in the third question, 4 in 100,000 in the fourth. Under standard assumptions, the third WTP should be approximately three times the fourth WTP. However, out of the 395 respondents who completed both questions, 178 (45.1%) gave the same answer to both questions and a further 37 actually stated a higher WTP for the smaller risk reduction. As a result, the mean value for preventing a 'statistical' injury S was computed as 2.5 times higher on the basis of the fourth question than on the basis of the third. Using median responses narrowed the disparity a little, but one figure was still double the other. Bearing in mind that these were the same respondents answering adjacent questions about the very same injury, there should ideally have been no significant difference between the two inferred values for preventing a statistical case of injury S.

Besides the concerns about insufficient sensitivity at the individual level, various features of the aggregate data were worrisome. First, the estimated values of preventing statistical injuries seemed high: for example, based on sample means, the best estimate of the value of preventing a single case of injury W was more than £750,000, which was higher than the value being used by DTp at that time to prevent a fatality. Second, the relative values seemed to have dubious implications: on the basis of sample means, preventing six injuries W (involving between them less than 2 years of somewhat reduced health and no further disability for the rest of life) was valued more highly than preventing a fatality (many of which involved between 20 and 50 years total loss of life).

The SG part of the study told a radically different story.

The basic set-up for each SG question involved placing one injury description card (just the injury description, not the risk reduction information used in the CV questions) on the left-hand side of a table top, with two cards on the right-hand side. One of these, set a little higher up the table, was always J ('Your normal state of health') while the other right-hand side card, set a little lower down, was a state mostly ranked worse than the one on the left – sometimes K (death), sometimes one of the other non-fatal scenarios.

In each question, the respondent was asked to suppose that they had been taken to hospital following a road accident and now had to choose between two ways of treating them. One option would – for certain – give the results shown on the single card on the left-hand side. The outcome of the other option was uncertain. If it succeeded, it would restore them quickly to their normal state of health – the upper of the two right-hand cards. If it failed, the outcome would be as shown on the lower of the two right-hand cards. The respondent was shown a list of 'chances of success' alongside a

list of the complementary ‘chances of failure’ and was asked to identify the success/failure combination of chances that would make them consider the certain and uncertain options to be equally balanced.

This form of question is consonant with EUT. If the two cards on the right-hand side are J and K and if normal health is assigned a utility of 1 and death is assigned a utility of 0, the utility of a health state/injury description I that lies between normal health and death is given as $u(I) = p \cdot 1 + (1-p) \cdot 0$, where p and $1-p$ are the chances of success and failure that make the respondent indifferent between the two options.

There were eight such SG questions in total, but the analysis in the TRL report focused on the three where the two right-hand side cards were J and K and where the left-hand cards were R, S and X, plus the one where W was on the left with J and S on the right.¹⁶ Under the assumptions of EUT, these questions allowed the values of preventing each of these four injuries to be calculated relative to the VoSL. These were very different from the relativities inferred from the CV sample’s responses to those WTP questions. For example, whereas the mean estimates from the CV questions had valued preventing a single injury W as about 18% of the value of preventing a fatality, the corresponding SG figure was 2%. And even this mean of 2% was arguably driven by a small number of high responses, since 81.5% of respondents declined to take *any* risk of death to avoid W.¹⁷

This study was important in the evolution of the research programme in two ways.

First, it demonstrated that CV questions of the kind used in this and earlier studies could not be relied upon to deliver values in which policymakers could have confidence. Asking respondents to make a more-or-less instantaneous¹⁸ series of separate trade-offs between sums of money and various small changes in already small probabilities of a spectrum of unfamiliar health states produced too many results that were difficult to reconcile with theory and/or with plausibility.

Second, the SG format offered an alternative possibility for eliciting people’s preferences in a way that was compatible with the underlying theory (EUT) but that seemed rather less demanding and appeared to yield more plausible results. Respondents seemed to find it easier and more meaningful to think about trade-offs within the domain of health states.

However, while SG questions could deliver the kinds of relativities which were the aim of this study, they could not by themselves identify an ‘absolute’ money value. So, given the doubts about the CV results in this particular study, the pragmatic decision

¹⁶It had been found in piloting that many respondents would not accept any risk of death when the left-hand ‘treatment’ would certainly deliver complete recovery within 4 months. There was greater willingness to take some risk of the more severe but non-fatal injury S, allowing a utility index number to be assigned to W such that $u(W) = q \cdot 1 + (1 - q) \cdot u(S)$, where q and $(1 - q)$ are the chances of success and failure producing indifference in this case.

¹⁷The relativity between W and K was obtained by an indirect route via S, so a ‘zero acceptable risk’ result could be produced either by refusing to take any risk of death in order to avoid S or else by refusing to take any risk of S to avoid W, or both. However, it would appear that the 81.5% is not overstated: for injury X, which entailed more time in hospital and considerably longer to recover, the question directly relating X to K produced a 75.2% refusal to take any risk of death.

¹⁸The average time taken to complete a CV survey was 45 min. Prior to the start of the interview, the respondents had no means of preparing for the kinds of judgements they were to be asked to make.

at the time was to take the existing VoSL (£620,000 in 1990 prices) and use the SG responses to scale all of the non-fatal injuries accordingly. But this was regarded as less than satisfactory. DTp had set the VoSL at £500,000 in 1987 prices as a 'consensus' figure based on the NOP study (Jones-Lee *et al.*, 1985) and various other empirical studies, combined with a helping of (cautious) civil service judgement (Dalvi, 1988), and had then adjusted that figure for inflation. Ideally, though, the appropriate monetary 'peg' would have been elicited from the same preference base as the relativities; but it was clear that the CV questions in this study were not robust enough to fulfil that role.

The 'chaining' method

When, in the later 1990s, a consortium of UK government departments and agencies commissioned further work to try to obtain a 'core' set of values for preventing deaths and injuries and also to explore whether values in contexts other than transport might be different, a team of economists and psychologists undertook several pilot studies and then a main study. Some of the results of piloting are reported in Beattie *et al.* (1998) which, despite making a number of fresh efforts to formulate 'direct' contingent valuation questions that would attenuate some of the earlier shortcomings, concluded that the results cast 'serious doubt on the reliability and validity of WTP based monetary values of safety estimated using conventional contingent valuation procedures' (Beattie *et al.*, 1998, Abstract).

The design of the main study, therefore, built on insights from the SG part of the 'non-fatals' study (Jones-Lee *et al.*, 1993), aiming 'to break the wealth/risk of death trade-off down into a number of conceptually manageable steps, thereby trying to attenuate the various biases that appear to be pervasive in responses to more direct contingent valuation questions in the health and safety field' (Carthy *et al.*, 1999, Abstract).

Briefly, the 'conceptually manageable steps' were as follows:

- (1) Obtain an estimate of the marginal rate of substitution (MRS) of wealth for an injury I whose impacts were reasonably easy for respondents to imagine and relate to. The estimate would be a weighted average of WTP(I) and WTA(I), where WTP(I) is the respondent's maximum WTP for a certain and very quick cure for I, and where WTA(I) is the minimum sum of money that would be just enough to offset experiencing I.
- (2) Use an SG formulation to elicit the utility loss of I relative to death.
- (3) 'Chain' (1) and (2) together to produce a value for preventing a fatality.

The main challenges, and how they were addressed, were as follows:

First, focusing WTP and WTA on the *certainty* of avoiding or experiencing an injury removed one of the practical problems exhibited in all of the studies described above, namely the difficulty of working with very small probabilities. However, this meant that the change could no longer be regarded as marginal.¹⁹ Hence, the need

¹⁹Even for the less serious injury W (broadly as described earlier), the median WTP was £1,000, which would have taken the average income earner a month to generate.

to deploy the theoretical apparatus to justify an estimate based on some weighted average of WTP and WTA (as well as the need to persuade respondents to take the WTP and WTA scenarios at face value).

Second, although the SG questions had been strongly favoured in the 1993 study, they had not been without drawbacks of their own. In particular, in that study, there were very many who said that they would not take *any* risk of death in return for being returned to normal health, even when the default prospect was a prognosis involving some permanent impairment. In case this was a form of ‘certainty effect’, the default scenario in the new study was modified so that even under the conventional treatment the respondent could not focus on a *zero* risk of a poor outcome. So in the SG question involving injury X (much the same as X in the previous section), the ‘conventional’ treatment involved a 999/1000 chance of X but also a 1/1000 chance of death. The respondent was then asked to identify what risk of death greater than or equal to 1/1000 they would accept in order to be restored to normal health if an alternative treatment were successful. It was thought – and this proved to be the case – that this scenario would reduce the numbers of respondents unwilling to consider any (additional) risk, thereby enabling this link in the chain to have a finite value for the great majority.

Third, it was felt that because of the unfamiliarity of the topic and the demanding nature of the tasks involved, it would help if all interviews were undertaken by members of the research team, whose deeper understanding of the issues would enable them to answer any questions or concerns respondents might have. The intention was to give the elicitation methods every chance to succeed.

In a number of respects, the results were encouraging. Without the complexity of information about small probabilities, answers to the WTP and WTA questions exhibited a good deal more sensitivity to the relative severity of injuries X and W: means and medians differed by factors of 3. For the SG questions, there was a much smaller proportion of respondents unwilling to take *any* extra risk in order to have a chance of avoiding X or W.

Even so, there were some features of the data that did not sit easily with standard assumptions: for example, for both X and W, some 18% of the sample gave WTA responses that were actually lower than their WTP responses, while at the other end of the spectrum another 20% of the sample gave WTA responses that were at least 10 times bigger than their WTP counterparts. At the individual level, the estimated weighted average could vary considerably depending upon which format of utility function was assumed. And different chaining sequences could produce substantially different VoSLs,²⁰ with some combinations of SG responses producing very large individual values that inflated the means very considerably.²¹

²⁰One (more direct) chain involved taking the estimated MRS for injury X and combining it with the relativity from the SG question linking X to death to produce a VoSL. The other (more indirect) chain took the MRS for W and linked it to the relativity produced by the combination of the two SG questions. The more indirect route gave VoSLs that were, on average, more than six times higher than their more direct counterparts. These mean differences were partly driven by some high values in a skewed distribution; but even the medians for the indirect chains were more than double their direct counterparts.

²¹To illustrate: if a respondent is only prepared to take an extra 1 in 1000 risk of death in a bid to cure X and if he/she is only prepared to take an extra 4 in 100 risk of X in the hope of avoiding W – neither of

Such results raised again a couple of related issues that Mike and his colleagues had wrestled with earlier (and were to confront again subsequently). First, on what grounds might one ‘trim’ individual observations that may appear to have a ‘disproportionate’ impact on mean values when means are regarded as the measures founded in the relevant theory? Or might the ‘median voter’ concept in political economy give some substantial weight to that measure of central tendency for public policy purposes?

In the event, Carthy *et al.* (1999) concluded in favour of using the mean values from the more direct chain (arguably less vulnerable to compounding errors) with a little trimming of a couple of extreme cases, and with a nod towards medians that were about one-third of those means. The resulting figure of £1 million was only a little higher than the figure already being used by the then Department of the Environment, Transport and the Regions: namely, £902,500 in 1997 prices.

Since then, the value assigned to preventing a statistical fatality has been periodically uprated in line with increases in UK prices generally and corresponding adjustments have been made to the values for preventing non-fatal injuries. At the time of writing, the VoSL used in road transport appraisal is in the region of £2 m while the figure for preventing a serious injury is a little over £200,000.

Joining up thinking across sectors and measurement units

While Mike’s primary focus had been upon the values of reducing risks of fatalities and injuries of various severities in the context of transport safety, a somewhat different stream of work had been more prominent in the field of health economics. There, much of the research had been less in the tradition of CBA and was more concerned with cost-effectiveness/cost-utility analysis. That is, taking a particular health care budget as given (by historical factors and the political process), how could that money be allocated in order to maximise the health-related benefits?

Health-related benefits are standardly measured using Quality Adjusted Life Years (QALYs). The quality of life in a given state of health is rated on a scale on which 0 corresponds with death and 1 with full health. A year of life at a state of health with a rating of h counts as h QALYs. The benefit of a proposed treatment (often but not always some new pharmaceutical product) which extends a patient’s life and/or improves their health state over some period of time is measured by the increase in the patient’s lifetime total of QALYs.

As the process is described above, no patient-preference-based monetary value is assigned to a QALY, so that no cost–benefit calculation is made. Rather, for many years in the UK, the National Institute for Health and Care Excellence (NICE) operated a (somewhat fuzzy) threshold system whereby those new treatments costing less than about £30,000 per QALY were likely to be recommended for adoption by the National Health Service while those costing much more than that were not.²²

which response is self-evidently implausible – and if the resulting relativity is chained to the median WTP for W (£1000), the end result is a VoSL of around £25 million in 1997 prices.

²²This is an oversimplification: over the years, various exceptions and special cases have been allowed; but the mainstream policy was broadly as described.

However, even though the transport safety and health evaluation fields had evolved somewhat separately, there was clearly scope for relating the two. Mike and colleagues explored that in two ways: in one direction, by considering how one might go from a VoSL to infer a monetary value of a QALY; in the other direction, by adapting the 'chaining' procedure to generate estimates of individual preference-based monetary values for QALYs.

The derivation of a value of a QALY from the VoSL involved Mike as theorist working with two Newcastle health economists (see Mason *et al.*, 2009). Depending on how (or whether) one adjusts for quality of health state and/or factors in some discount rate to allow for supposed pure time preference, the authors' Table II illustrates the range and variability of possible values. Taking the VoSL (at 2005 prices) as £1.4 m and the average number of years of remaining life expectancy lost in road traffic fatalities to be 32 years, an undiscounted non-quality adjusted life year was valued at £23,200 while a discounted QALY was valued at £70,900. So the study demonstrated the ways in which VoSLs, QALYs and statistical life years (SLYs) could be related to one another while at the same time drawing attention to the dependence of the estimates upon variables for which the empirical basis was limited.²³

The Mason *et al.* (2009) paper was one output from a larger group of projects commissioned by the (then) National Coordinating Centre for Research Methodology (NCCRM), most of which were devoted to exploring different stated preference methods for eliciting weights and values for QALYs. The broader set of empirical studies and the preliminary results obtained from them are reported in Baker *et al.* (2010). The one that was most pertinent to Mike's research is described in Chapter 6 of that report, under the heading 'Exploring the feasibility of eliciting the monetary value of a QALY' (pp. 47–63). This provided another opportunity to test components of the 'chaining' method.

The essential features of the study were as follows. Two types of illness were described in 'natural' language. For each type, there were three duration scenarios: the condition lasted for 3 months followed by a return to normal health; it lasted for 12 months, after which normal health returned; or it was a lifelong condition. Various WTP questions were asked, including for the avoidance of the certainty of some scenarios and for the elimination of a specified risk (respondents were randomised between 5% and 10% risks) of some outcomes. Standard gamble questions examined the relativities of different durations and also sought to elicit individual indices for the quality of the health state under consideration.

As in earlier studies reviewed above, divergences between theoretical expectations and empirical regularities were in evidence. Many of the WTP and SG distributions were heavily skewed, with the mean WTP responses being insufficiently sensitive to the different risks and durations. At the same time, the SG responses were difficult to reconcile with standard assumptions about time preferences. Different patterns of

²³For example, not having an estimate of a discount rate based on an appropriate empirical study, the authors used the 1.5% figure recommended by the UK Treasury at that time; and quality of life weights were taken from a working paper (Kind *et al.*, 1999). No claims were made by Mason *et al.* (2009) for the validity of those numbers.

chaining produced different values of a QALY, and were prone to generate individual-level values that appeared far too high: for example, combining WTP to avoid a 12-month illness with the health state utility index produced by the appropriate SG question resulted in more than a quarter of the sample being estimated to have a value of a QALY of more than £1 million – i.e. more than 30 times the figure used by NICE. Whatever the attractions of the chaining approach in theory, and however much had been learned about survey design during nearly three decades of trying to elicit preferences for health and safety, the gap between the assumptions of the modelling and the patterns in the data persisted.

Finale

Officially, Mike retired to an emeritus professorship at Newcastle University in 2009, but he remained committed to research into individuals' valuations of risk reductions. In his supposed retirement, he published a further 20 or so research papers on this and related topics, to add to the more than a hundred that had preceded them. When, in 2017, a consortium of UK government departments and agencies sought to commission research to provide fresh estimates of VoSLs and their relationship with values of SLYs and QALYs, Mike saw an opportunity to pull together many threads of his previous theoretical and applied work. He was one of a group of collaborators from Newcastle, Glasgow Caledonian and Birmingham universities who succeeded in a bid to undertake a 'scoping study' of the relevant literature (Chilton *et al.*, 2020). This study produced a set of recommendations for a new round of empirical work, underpinned (as Mike's work had always been) by EUT. However, the advent of COVID-19 in 2020 put the commissioning process on hold. Sadly, Mike's health deteriorated later that year and he died before the process could restart.

The messages recorded in Newcastle University's Book of Condolences²⁴ express universal agreement about Mike's qualities as an academic. He was a brilliant thinker and a superb lecturer, totally in command of the rigorous formal theory on which he worked, but able to make it seem simple to everyone else. (One student from long ago still remembered Mike's use of a dustbin lid in a lecture on constrained optimisation.) He was unpretentious, generous with his time, but equipped with what one contributor described as a 'bullshit radar'. The fundamental seriousness and intellectual honesty of his research was leavened with a mischievous and irreverent but always good-natured sense of humour.

Achievements

The project of finding stated-preference-based monetary values of statistical life, safety and health is now a well-established research programme. Mike has a strong claim to be regarded as one of its founders and arguably its most determined and tenacious advocate during the past five decades. Of the small group of early proponents of the idea of valuing statistical life, it was Mike who took the first serious steps to use representative stated preference surveys to find values that policymakers

²⁴<https://blogs.ncl.ac.uk/bookofcondolence/>, posted by Newcastle University Business School.

could actually use. Over the following years, Mike was the intellectual leader of a growing group of researchers in the field, many of whom began their careers with Mike as their mentor.²⁵ In this final section, we consider what this programme has achieved, and how it relates to wider developments in economics.

The fundamental normative premise of the programme is that, when public policy decisions require judgements about the monetary value of safety improvements, those judgements should be aligned with the considered preferences of informed citizens. Empirically, it is assumed that these preferences can be elicited directly in surveys and that, to a reasonable degree of approximation, they have the consistency properties of EUT. We have described Mike's work as a progression of efforts to increase the fit between theory and survey responses – partly by refining the theory, but primarily by improving survey designs.

This strategy treats the conditions that define a person's 'true' preferences, namely adequate information and careful reflection, as standards to be achieved by the survey design that is used to elicit them; and it treats finding a satisfactory survey design as one of its research problems. 'Anomalous' survey responses – that is, responses that are inconsistent with EUT – are not brushed aside as evidence of respondents' tendencies to error. Instead, they are brought to light and seen as pointers to possible design improvements. The hope is that, by following this strategy, research will progressively reduce the extent of anomalies. That hope rests on a belief that the EUT axioms capture broad structural properties of people's considered preferences. Mike kept that hope and that belief.

At this point, we (Graham and Bob) should say that we do not entirely share Mike's faith in this respect. When we first worked in experimental economics, we too expected to find that individuals' choices would be broadly consistent with some theory of rational choice, although we were more ready than Mike to consider theories that contravened EUT axioms. We (and particularly Graham, often working with Mike) investigated methods of eliciting individuals' supposedly consistent preferences to provide guidance to policymakers. Over time, we have both come to doubt whether such preferences really exist. Here, however, it is we and not Mike who are deviating from mainstream opinion in economics.

From the early 2000s, many other behavioural economists have proposed approaches which assume that individuals have 'true' preferences with standard rationality properties but that, when making actual choices, those individuals are susceptible to 'bias' and 'error'. Normative analysis is based on true preferences, which the economist reconstructs by removing the supposed effects of bias and error from the preferences revealed in actual choices. Mike's research strategy is similar in using a concept of true preference, but it does so in a braver way. By defining true preferences in terms of conditions of good survey design, it accepts responsibility for finding an operational method of identifying those preferences – a responsibility that is evaded in many behavioural analyses of bias and error.²⁶

²⁵The 'References' section lists papers which Mike co-authored with a total of 38 different collaborators.

²⁶This critique is advanced by Infante *et al.* (2016) in a review of contributions to 'behavioural welfare economics'.

However, it would be wrong to suggest that the success of Mike's research programme depends entirely on how well an EUT model can be fitted to survey responses. For all Mike's commitment to EUT, there was a pragmatic strand in his research. In reporting the results of stated preference studies, he recognised that policymakers needed to set specific monetary values for reductions in specific types of risk. While being open about anomalies in the survey responses he reported, Mike had to advise policymakers on the basis of currently available data. Typically, he argued that these data were good enough to identify plausible ranges of values. For example, the conclusion of the first DTp study presented the results as supporting 'the inference of at least the broad order of magnitude of the VoSL for transport risks', specifically that, in 1987 prices, this should be 'at a minimum £600,000 and more probably well in excess of £1,000,000' (Jones-Lee, 1989: 206). Mike was more precise about the minimum than the maximum because even the lowest values in the plausible range were substantially larger than those used by the DTp at the time – recall that, despite the results of the NOP study, the DTp had opted to set the figure at £500,000 in 1987 prices.

From at least as far back as his first 'illustrative survey', Mike believed that UK government departments undervalued safety. (In *The Value of Life*, he describes the findings of that survey as 'prima facie evidence of a severe underexpenditure on safety and longevity-improvement' (Jones-Lee, 1976: 150).) Realistically, he accepted that his conclusions about the correct VoSL – the value that best represented individuals' preferences – were too radical for most policymakers to accept. In consequence, the preference-based justifications for the VoSL increases that actually resulted from Mike's work were very robust. Mike rightly took pride in his contribution to making governments more willing to spend on safety.

References

- Baker, R., I. Bateman, C. Donaldson, M. Jones-Lee, E. Lancsar, G. Loomes, H. Mason, M. Odejar, J. L. Pinto Prades, A. Robinson, M. Ryan, P. Shackley, R. Smith, R. Sugden and J. Wildman (2010), 'Weighting and valuing quality-adjusted life-years using stated preference methods: preliminary results from the Social Value of a QALY Project', *Health Technology Assessment*, **14**(27): 1–157.
- Beattie, J., J. Covey, P. Dolan, L. Hopkins, M. Jones-Lee, G. Loomes, N. Pidgeon, A. Robinson and A. Spencer (1998), 'The contingent valuation of safety and the safety of contingent valuation, part 1: caveat investigator', *Journal of Risk and Uncertainty*, **17**: 5–25.
- Broome, J. (1978), 'Trying to value a life', *Journal of Public Economics*, **9**: 91–100.
- Broome, J. (1985), 'The economic value of life', *Economica*, **52**: 281–294.
- Carthy, T., S. Chilton, J. Covey, L. Hopkins, M. Jones-Lee, G. Loomes, N. Pidgeon and A. Spencer (1999), 'The contingent valuation of safety and the safety of contingent valuation, part 2: the CV/SG 'chained' approach', *Journal of Risk and Uncertainty*, **17**: 187–213.
- Chilton, S., M. Jones-Lee, H. Metcalf, J. S. Nielsen, R. Baker, C. Donaldson, H. Mason, N. McHugh, R. McDonald and M. Spackman (2020), A scoping study on the valuation of risks to life and health: the monetary value of a life year (VOLY). Retrieved from: <https://www.gov.uk/government/publications/valuation-of-risks-to-life-and-health-monetary-value-of-a-life-year-voly/a-scoping-study-on-the-valuation-of-risks-to-life-and-health-the-monetary-value-of-a-life-year-voly>.
- Dalvi, M. Q. (1988), *The Value of Life and Safety: The Search for a Consensus Estimate*. London: Department of Transport.
- Dasgupta, P., S. Marglin and A. Sen (1972), *Guidelines for Project Evaluation*. New York: United Nations Industrial Development Organization.

- Dawson, R. F. F. (1967), *Costs of Road Accidents in Great Britain*. London: Road Research Laboratory, Ministry of Transport.
- Dawson, R. F. F. (1971), *Current Costs of Road Accidents in Great Britain*. London: Road Research Laboratory, Department of the Environment.
- Dean, J. (1951a), *Capital Budgeting*. New York: Columbia University Press.
- Dean, J. (1951b), *Managerial Economics*. New York: Prentice-Hall.
- Dolan, P., M. Jones-Lee and G. Loomes (1995), 'Risk-risk versus standard gamble procedures for measuring health state utilities', *Applied Economics*, **27**: 1103–11.
- Drèze, J. (1962), 'L'utilité sociale d'une vie humaine', *Revue Française de Recherche Opérationnelle*, **23**: 93–118.
- Dubourg, R., M. Jones-Lee and G. Loomes (1994), 'Imprecise preferences and the WTP-WTA disparity', *Journal of Risk and Uncertainty*, **9**: 115–133.
- Hammerton, M., M. Jones-Lee and V. Abbott (1982a), 'The consistency and coherence of attitudes to physical risk', *Journal of Transport Economics and Policy*, **16**: 181–199.
- Hammerton, M., M. Jones-Lee and V. Abbott (1982b), 'Equity and public risk: some empirical results', *Operations Research*, **30**: 203–207.
- Infante, G., G. Lecouteux and R. Sugden (2016), 'Preference purification and the inner rational agent: a critique of the conventional wisdom of behavioural welfare economics', *Journal of Economic Methodology*, **23**: 1–25.
- Jones-Lee, M. (1969), 'Valuation of reduction in probability of death by road accident', *Journal of Transport Economics and Policy*, **3**(1): 37–47.
- Jones-Lee, M. (1971), 'Selected Problems in the Theory of Economic Behaviour under Risk and Uncertainty', *D.Phil. Dissertation*, University of York.
- Jones-Lee, M. (1974), 'The value of changes in the probability of death or injury', *Journal of Political Economy*, **82**(4): 835–849.
- Jones-Lee, M. (1976), *The Value of Life: An Economic Analysis*. London: Martin Robertson.
- Jones-Lee, M. (1979), 'Trying to value a life: why Broome does not sweep clean', *Journal of Public Economics*, **12**(2): 249–256.
- Jones-Lee, M. (1987), 'The economic value of life: a comment', *Economica*, **54**: 397–400.
- Jones-Lee, M. (1989), *The Economics of Safety and Physical Risk*. Oxford: Blackwell.
- Jones-Lee, M., M. Hammerton and P. Philips (1985), 'The value of safety: results of a national sample survey', *Economic Journal*, **95**: 49–72.
- Jones-Lee, M., G. Loomes, D. O'Reilly and P. Philips (1993), 'The value of preventing non-fatal road injuries: findings of a willingness-to-pay national sample survey. TRL Contractor Report 330. Retrieved from: <https://trid.trb.org/view/377578>.
- Kind, P., G. Hardman and S. Macran (1999), *UK Population Norms for EQ-5D. Discussion Paper 172, Centre for Health Economics*. UK: University of York. Retrieved from: <https://ideas.repec.org/p/chy/respap/172chedp.html>.
- Little, I. and J. Mirrlees (1974), *Project Appraisal and Planning for Developing Countries*. London: Heinemann.
- Mason, H., M. Jones-Lee and C. Donaldson (2009), 'Modelling the monetary value of a QALY: a new approach based on UK data', *Health Economics*, **18**: 933–950.
- Melinek, S. (1974), 'A method of evaluating human life for economic purposes', *Accident Analysis and Prevention*, **6**: 103–114.
- Mishan, E. (1971a), *Cost-Benefit Analysis*. London: Allen and Unwin.
- Mishan, E. (1971b), 'Evaluation of life and limb: a theoretical approach', *Journal of Political Economy*, **79**(4): 687–705.
- O'Reilly, D., J. Hopkin, G. Loomes, M. Jones-Lee, P. Philips, K. McMahon, D. Ives, B. Soby, D. Ball and R. Kemp (1994), 'The value of road safety: UK research on the valuation of preventing non-fatal injuries', *Journal of Transport Economics and Policy*, **28**: 45–59.
- Reynolds, D. J. (1956), 'The cost of road accidents', *Journal of the Royal Statistical Society*, **119**(4): 393–408.
- Schelling, T. (1968), 'The life you save may be your own', in S. B. Chase (ed.), *Problems in Public Expenditure Analysis*, Washington, DC: Brookings Institution, 127–162.
- Slovic, P., B. Fischhoff and S. Lichtenstein (1981), 'Perceived risk: psychological factors and social implications', *Proceedings of the Royal Society A*, **376**: 17–34.

- Smith, R. (1973), Compensating wage differentials and hazardous work. Technical Paper No. 5. Office of Policy Evaluation and Research, US Department of Labor.
- Sugden, R. and A. Williams (1976), *The Principles of Practical Cost-Benefit Analysis*. Oxford: Oxford University Press.
- Thaler, R. and S. Rosen (1973), *The Value of Saving a Life: Evidence from the Labor Market*. University of Rochester Graduate School of Management. Working Paper No. 7401.
- Williams, A. (1972), 'Cost-benefit analysis: bastard science and/or insidious poison in the body politick?' *Journal of Public Economics*, **1**: 199–225.